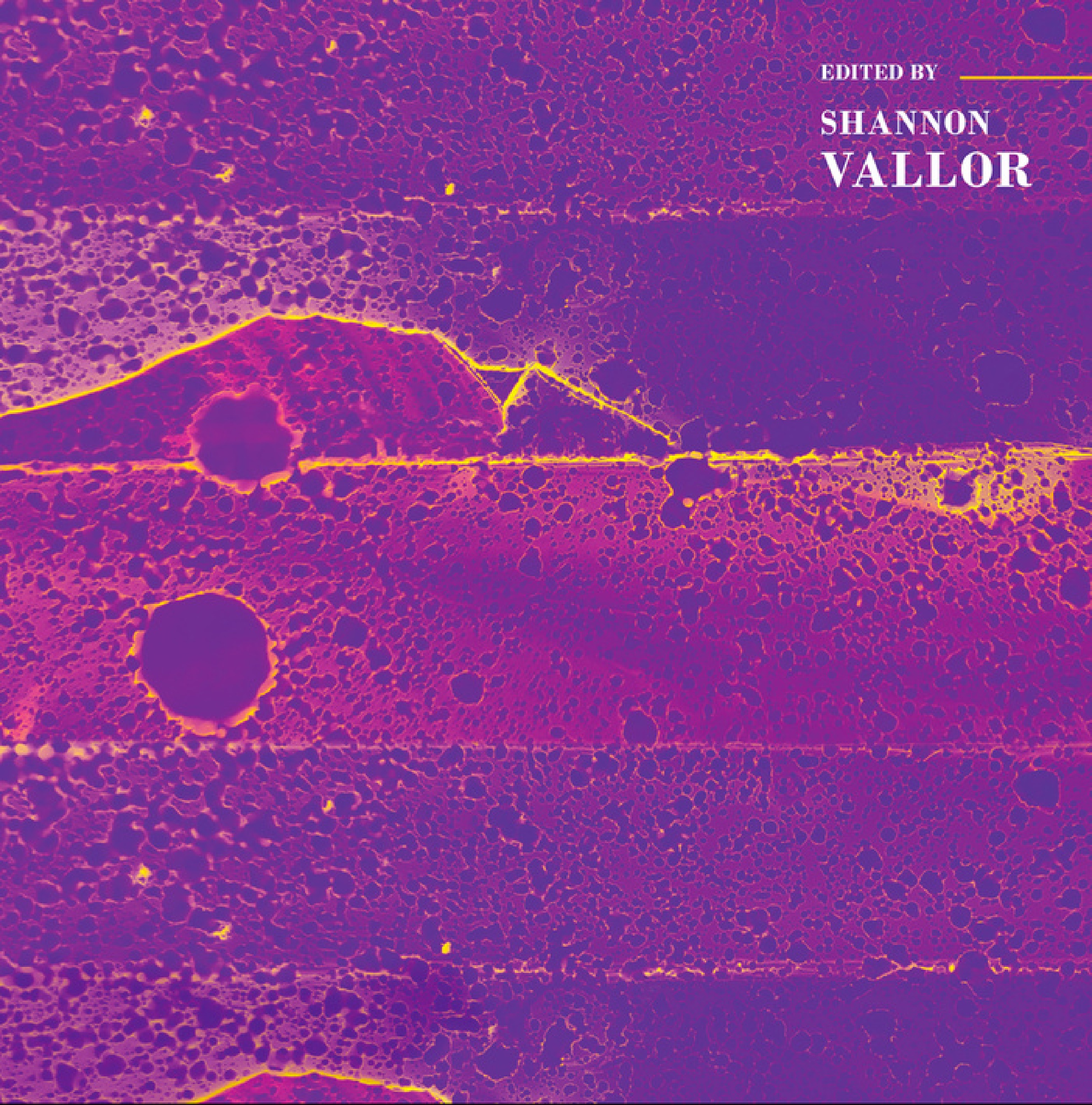

EDITED BY

SHANNON
VALLOR

The Oxford Handbook *of* PHILOSOPHY OF TECHNOLOGY

THE OXFORD HANDBOOK OF

PHILOSOPHY OF TECHNOLOGY

THE OXFORD HANDBOOK OF PHILOSOPHY OF TECHNOLOGY

Edited by

SHANNON VALLOR

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

Library of Congress Control Number: 2021039934

ISBN 978–0–19–085118–7

DOI: 10.1093/oxfordhb/9780190851187.001.0001

1 3 5 7 9 8 6 4 2

Printed by Sheridan Books, Inc., United States of America

Contents

PART I: HISTORIES AND METHODOLOGIES IN THE PHILOSOPHY OF TECHNOLOGY

PART II: TECHNOLOGY AND EPISTEMOLOGY

PART III: TECHNOLOGY, POWER, AND POLITICS

PART IV: TECHNOLOGY, METAPHYSICS, AND LANGUAGE

PART V: TECHNOLOGY, AESTHETICS, AND DESIGN

PART VI: TECHNOLOGY, HEALTH AND THE ENVIRONMENT

PART VII: TECHNOLOGY AND THE GOOD LIFE

Author Biographies

Keith Abney is Senior Research Fellow of the Ethics + Emerging Sciences Group at California Polytechnic State University, San Luis Obispo, and senior lecturer in the Philosophy Department. He is co-editor of *Robot Ethics* (MIT Press) and *Robot Ethics 2.0* (OUP) as well as author/contributor to numerous other books, journal articles, and grant-funded reports for the National Science Foundation, the Greenwall Foundation, and the US Office of Naval Research. His current research ranges through many aspects of emerging technology ethics and bioethics, from co-editing a special journal edition on the ethics of human colonization of other planets to work on military AI, space war, and cyberethics to problems of abuse in autonomous vehicles, the ethics of human enhancements, and more.

Alison Adam is Professor Emerita of Science, Technology, and Society at Sheffield Hallam University, UK. She is the author of *Artificial Knowing: Gender and the Thinking Machine* (Routledge, 1998); *Gender, Ethics and Information Technology* (Palgrave, 2005); and *A History of Forensic Science: British Beginnings in the Twentieth Century* (Routledge, 2016). She is co-editor of *Women, Work, and Computerization: Breaking Old Boundaries—Building New Forms* (Elsevier, 1994); *Women in Computing: Progress from Where to What? (*Intellect,1997); *Virtual Gender: Technology, Consumption, and Identity (*Routledge, 2001); and editor of *Crime and the Construction of Forensic Objectivity from 1850*, (Palgrave, 2020). Her current research focuses on the construction of forensic objectivity in forensic science and is centered on mid-twentieth century Scotland.

Ciano Aydin is Professor of Philosophy of Technology, Head of the Department of Philosophy, and Vice-Dean of the Faculty of Behavioural, Management, and Social Sciences at the University of Twente, The Netherlands. His research focuses on "existential technology"; he investigates how technologies increasingly shape our identity, impact our freedom and responsibility, and influence different facets of our life. His main areas of interest are philosophy of technology, philosophy of self, phenomenology, philosophy of mind, and ethics. He has published in *Transactions of the Charles S. Peirce Society*, *Phenomenology and the Cognitive Sciences*, *Journal of Medicine and Philosophy*, *Philosophy and Technology,* and other journals. In 2021 he published the book *Extimate Technology: Self-Formation in a Technological World* (Routledge). See www.cianoaydin.nl for more information about his current research.

Philip Brey is Professor of Philosophy and Ethics of Technology at the Department of Philosophy, University of Twente, the Netherlands. He is a former president of the

Society for Philosophy and Technology (SPT) and of the International Society for Ethics and Information Technology (INSEIT). His research focuses on ethics and philosophy of emerging technologies, in particular digital technologies, AI, robotics, biomedical technology, and sustainable technology. Part of his work has focused on developing novel approaches in the philosophy and ethics of technology, including anticipatory technology ethics, ethical impact assessment, structural ethics, a novel extension theory of technology, and ethics by design for artificial intelligence. He currently coordinates the 10-year research program *Ethics of Socially Disruptive Technologies* that includes seven universities in the Netherlands and over 50 researchers.

Adam Briggle is Associate Professor and the Director of Graduate Studies in the Department of Philosophy and Religion at the University of North Texas. His publications include *A Rich Bioethics: Public Policy, Biotechnology, and the Kass Council* (Notre Dame, 2010); *A Field Philosopher's Guide to Fracking* (Liveright, 2015); and *Thinking through Climate Change: A Philosophy of Energy in the Anthropocene* (Palgrave Macmillan, 2021). His research interests are in public philosophy with a focus on energy and the environment. He is currently working on developing effective pedagogical strategies for teaching the ethics and politics of climate change.

Sage Cammers-Goodwin is a PhD Candidate at the University of Twente in Enschede, The Netherlands. She completed her prior studies in Computer Science and Symbolic Systems at Stanford University. Her work centers on tech ethics and the responsibility corporations and governments hold in an increasingly digital world. Her master's thesis "*Tech: The Curse and the Cure*" was published in UCI Law Review and explores misconceptions about tech culture being a common good. Currently, her research centers on smart cities and the policies that will be needed to support their ethical development.

Mark Coeckelbergh is Professor of Philosophy of Media and Technology and Vice Dean of the Faculty of Philosophy and Education at the University of Vienna. He is a philosopher of technology and an expert in the ethics of robotics, AI, and (other) automation technologies. Previously he was President of the Society for Philosophy and Technology, Professor of Technology and Social Responsibility at De Montfort University, and Managing Director of the 3TU Centre for Ethics and Technology in the Netherlands. He is the author of numerous publications, including *Using Words and Things* (Routledge, 2017); *Introduction to Philosophy of Technology* (Oxford University Press, 2019); *AI Ethics* (MIT Press, 2020); and *Narrative and Technology Ethics* (Palgrave, 2020), with Wessel Reijers. Currently he works on the political philosophy of artificial intelligence in a global context.

John Danaher is a Senior Lecturer in the School of Law, NUI Galway, Ireland. He is the co-author of *A Citizen's Guide to Artificial Intelligence* (MIT Press 2021), the author of *Automation and Utopia: Human Flourishing in a World Without Work* (Harvard University Press, 2019), and the co-editor of *Robot Sex: Social and Ethical Implications* (MIT Press 2017). He has published several dozen academic papers on different topics, including the risks of advanced AI, the ethics of social robotics, the meaning of life and the future of work, and the ethics of human enhancement. His work has appeared in *The Atlantic, VICE: Motherboard, The Guardian, The Irish Times, The Sunday Times, Aeon, and The Philosophers' Magazine*.

Massimo Durante is Professor in Philosophy of Law and Legal Informatics at the Department of Law, University of Turin. He holds a PhD in Philosophy of Law, Department of Law, University of Turin, and a PhD in Moral Philosophy, Faculty of Philosophy, Paris IV Sorbonne. He is a Member of the Board of the Joint International Doctoral (PhD) degree in "Law, Science, and Technology" and of the Doctoral Program "Right of Internet of Everything." He is Faculty Fellow of the Nexa Center for Internet and Society at the Politecnico of Turin, and member of the Scientific Board of the Master in Data Protection Law at the Department of Law, University of Turin. Author of several books, he has widely published papers in Italian, English, and French. His main interests are law and technology, information ethics, digital governance and information technology law, privacy and data protection law, AI & law.

Charles Ess is Professor Emeritus, Department of Media and Communication, University of Oslo, Norway. He works across the intersections of philosophy, computing, applied ethics, comparative philosophy and religious studies, and media studies, with emphases on internet research ethics, digital religion, virtue ethics, social robots, and AI. Ess has published extensively on ethical pluralism, culturally variable ethical norms and communicative preferences in cross-cultural approaches to Information and Computing Ethics, and their applications to everyday digital media technologies; his *Digital Media Ethics*, 3rd edition, was published in early 2020. His current work concerns meta-theoretical and meta-disciplinary complementarities between ethics and the social sciences, applied ethics in ICT design and implementation (especially social robots and AI), and research ethics in Human-Machine Communication.

Maarten Franssen is Associate Professor of Philosophy at the Delft University of Technology. After obtaining degrees in physics and in history, he received a doctorate in Philosophy from the University of Amsterdam in 1997 for a thesis on foundational studies in the social sciences. Since coming to Delft, he has worked on various topics in analytic philosophy of technology: the epistemology and methodology of engineering, the ontology of artifacts, the various roles of normativity in technology, and sociotechnical systems. He is co-author of the book *A Philosophy of Technology: From Technical Artefacts to Sociotechnical Systems* (Morgan & Claypool 2011) and of the lemma on Philosophy of Technology in the *Stanford Encyclopedia of Philosophy*. He is co-editor of the volumes *Artefact Kinds: The Ontology of the Human-made World* (Springer Synthese Library 2014) and *Philosophy of Technology after the Empirical Turn* (Springer 2016).

Barbro Fröding is Docent and Senior Researcher at KTH Royal Institute of Technology Stockholm, Sweden. Fröding is a Former fellow at Lincoln College Oxford and was awarded a Merit Award by the Oxford Philosophy Faculty in the Humanities Division

Merit Award Exercise 2008/2009. Her main research interests include ethics, bioethics, virtue ethics, ethical aspects of medical technology, and the ethics of cognitive enhancement. She has also written about good decision-making concerning the development and use of technology with an eye to social sustainability: *Neuroenhancement: how mental training and meditation can promote epistemic virtue* (co-authored with Walter Osika, Springer 2015) and *Virtue Ethics and Human Enhancement* (Springer 2013). Currently she is in charge of a research project on the ethical aspects attaching to AI and data collection in the smart city and she is involved in an EU-funded research project on disruptive technologies in the public sector.

Julia D. Gibson is a Postdoctoral Fellow with the APPLE (Animals in Philosophy, Politics, Law, and Ethics) consortium at Queen's University. They envision their research taking shape where the boundaries between feminist, political, and environmental philosophy grow pleasantly and productively murky. Her publications include articles in *Environmental Philosophy*, the *Journal of Agricultural and Environmental Ethics*, and the *International Journal of Feminist Approaches to Bioethics.* Their research exploring the commonalities, tensions, and incommensurabilities of decolonial and interspecies justice finds relational and material expression in her public philosophy work on the family farm. Julia currently serves on the Executive Committee of the International Association for Environmental Philosophy.

Anna Gotlib is Associate Professor of Philosophy at Brooklyn College CUNY, specializing in feminist bioethics/medical ethics, moral psychology, and philosophy of law. She received her PhD in philosophy from Michigan State University, and a JD from Cornell Law School. Anna co-edits the *International Journal of Feminist Approaches to Bioethics.* Her work appeared in *The Kennedy Institute of Ethics Journal, Journal of Bioethical Inquiry, Journal of Medical Humanities, Hypatia, Aeon/Psyche,* and other publications. She also edited two volumes on moral psychology for Rowman and Littlefield International: *The Moral Psychology of Sadness* (2017) and *The Moral Psychology of Regret* (2019). In 2020, Anna was a Fulbright Specialist Scholar at the University of Iceland.

A. S. Aurora Hoel is Professor of Media Studies and Visual Culture at the Norwegian University of Science and Technology (NTNU). She is the recipient of several research fellowships and awards, most recently a Novo Nordisk Foundation Visiting Professorship in Art & Art History, focusing on machine images (Aarhus University, 2020); and a European Commission Horizon 2020 Marie Skłodowska-Curie Actions Individual Fellowship, focusing on image-guided surgery (Humboldt University, 2015–2017). Previously, she was the PI of the multidisciplinary research project *Picturing the Brain: Perspectives on Neuroimaging*, founded by the Research Council of Norway (NTNU, 2010–2014). Her current research probes deeper into the mediating roles of bodies and machines in knowledge and being, seeking to establish an operational aesthetics. Hoel has published widely in the overlapping fields of media theory, science studies, and the philosophy of technology.

Anna Lauren Hoffmann is Assistant Professor at the Information School of the University of Washington, an affiliate faculty member with the UW DataLab, and a senior fellow with the Center for Applied Transgender Studies. Her work focuses on issues in information, data, and ethics, with specific attention to the ways discourses and uses of data science and technology (and their purported ethics) inflict cultural and other violences and reinforce patterns of dominance and subjugation—especially along the lines of race, gender, and ability. Her writing on data, technology, and ethics has been published in *New Media & Society*, *Information, Communication, & Society*, *The Library Quarterly*, and the *Journal of the Association for Information Science and Technology*. She has also been awarded funding from the National Science Foundation for her research on data science ethics education. She is also an engaged public scholar and her writing has appeared in numerous venues, including *The Guardian*, *Slate*, and *The Los Angeles Review of Books*.

Wybo Houkes is Professor of Philosophy of Science and Technology at the Eindhoven University of Technology, the Netherlands. In 2008, he received a personal research grant from the Netherlands Organization of Scientific Research (NWO) for a project on evolutionary theories of technology. His publications include *Technical Functions* (Springer, 2010; with Pieter E. Vermaas) and numerous papers on technical artifacts, engineering knowledge, scientific modeling practices, and cultural-evolutionary theory. His current research includes a study of conditions for successful model transfer between research fields; an analysis of the negative effects of competition between researchers; and the effects of digitization and "servitization" on our involvement with technologies and basic consumer rights.

Don Howard is Professor of Philosophy at the University of Notre Dame, past director of Notre Dame's Reilly Center for Science, Technology, and Values, and an affiliate of the new Notre Dame Center for Technology Ethics. He has been writing and teaching about the ethics of science and technology for many years. He was co-editor of the collection, *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited* (University of Pittsburgh Press, 2008). His paper, "Virtue in Cyberconflict," appeared in *Ethics of Information Warfare* (Springer, 2014), his essay on "Civic Virtue and Cybersecurity" in *The Nature of Peace and the Morality of Armed Conflict* (Palgrave Macmillan, 2017), and his paper, "The Moral Imperative of Green Nuclear Energy Production" in the *Notre Dame Journal on Emerging Technologies* (2020). His editorials on technology ethics have appeared in the *Wall Street Journal*, on CNN, at *InsideSources*, *NBC Think*, and in other venues.

Deborah G. Johnson is the Olsson Professor of Applied Ethics, Emeritus at the University of Virginia in Charlottesville, Virginia. She received a Lifetime Achievement Award from the Society for Philosophy and Technology in 2021, the Covey Award from the International Association for Computing and Philosophy in 2018, and the Joseph Weizenbaum Award for life-long contributions to information and computer ethics from the INSEIT in 2015. Over the course of her career, Johnson has published seven books, including one of the first textbooks on computer ethics in 1985 [*Computer Ethics*, Prentice Hall] and most recently *Engineering Ethics, Contemporary and Enduring Debates* (Yale University Press, 2020). Her research addresses ethical, social, and policy implications of technology and engineering with her latest focus on issues around AI, robots, and deepfakes.

Sanna Lehtinen (PhD 2015, University of Helsinki) is a Research Fellow in the Transdisciplinary Arts Studies unit at the Aalto University School of Arts, Design and Architecture. Her research focuses on urban and environmental aesthetics and the philosophy of the city. In 2013 she was awarded the Young Scholar Award of the International Association of Aesthetics. Her publications include journal articles published in *Open Philosophy*, *Philosophical Inquiries*, *Essays in Philosophy*, and *Behaviour & Information Technology*, articles in edited volumes published by Routledge, Springer, and Oxford University Press, as well as editing special issues for *Contemporary Aesthetics* and *Open Philosophy*. She is the President of the Finnish Society for Aesthetics and Co-director on the board of the international Philosophy of the City Research Group. Her current research interests include the aesthetics of new urban technologies and environmental aesthetics in the era of climate change.

Anthonie Meijers is emeritus Professor of Philosophy and Ethics of Technology at the Eindhoven University of Technology, the Netherlands. He is the editor in chief of *Philosophy of Technology and Engineering Sciences*, Volume 9 of *Handbook of The Philosophy of Science*, Amsterdam: Elsevier, 2009, 1453 pp.). His publications include Kroes, P.A. and A.W.M. Meijers (2000, eds.), *The Empirical Turn in the Philosophy of Technology*, Research in Philosophy and Technology Vol. 20 (Elsevier Science), Amsterdam, 253 pp., and Kroes, P. A. and A. W. M. Meijers (eds.) (2006) *The Dual Nature of Technical Artefacts*, special issue of *Studies in History and Philosophy of Science* 37, 1–158. His current research focuses on the ethics of socially disruptive technologies.

Carl Mitcham is International Professor of Philosophy of Technology at the Renmin University of China and Emeritus Professor of Humanities, Arts, and Social Sciences at the Colorado School of Mines. Publications include *Thinking through Technology: The Path between Engineering and Philosophy* (1994), *Encyclopedia of Science, Technology, and Ethics* (4 vols., 2005), *Ethics and Science: An Introduction* (2012, with Adam Briggle), and *Steps toward a Philosophy of Engineering: Historico-Philosophical and Critical Essays* (2020). He has served on the Committee on Scientific Freedom and Responsibility of the American Association for the Advancement of Science (1994–2000) and expert study groups for the European Commission (2009 and 2012). Awards include the International World Technology Network Award for Ethics (2006), a Doctorate Honoris Causa from the Universitat Internacional Valenciana, Spain (2010), and a Lifetime Achievement Award from the Society for Philosophy and Technology (2021).

Samantha Noll is Assistant Professor in The School of Politics, Philosophy, and Public Affairs at Washington State University. She is also the bioethicist affiliated with the Functional Genomics Initiative, which applies genome editing in agriculture research, and the Center for Reproductive Biology. Dr. Noll is the co-author or editor of two books, including a *Field Guide to Formal Logic* (Great River Learning, 2020) and the *Routledge Handbook of Philosophy of the City* (Routledge, 2019). She publishes widely on food justice and food sovereignty, local food movements, and the application of biotechnologies in food production. She is also the author or co-author of more than 30 other publications, in journals ranging from *Environmental Ethics* to *Pragmatism Today*. She is currently working with scholars from several disciplines on various projects that engage with ethical considerations at the intersection of philosophy of food, environmental ethics, and emerging technologies.

Beth Preston is Professor Emerita of Philosophy at the University of Georgia in Athens, Georgia (United States). Her publications include *A Philosophy of Material Culture: Action, Function, and Mind* (Routledge, 2013), the "Artifact" entry in the *Stanford Encyclopedia of Philosophy,* and articles and book chapters on the epistemology and metaphysics of artifacts and their functions, such as "Why Is a Wing Like a Spoon? A Pluralist Theory of Function" (*The Journal of Philosophy*, 1998) and "Ethnotechnology: A Manifesto" (in *Artefact Kinds: Ontology and the Human-Made World,* Springer, 2014). Her current research focuses on the improvised nature of everyday activity in our familiar world of artifacts, people, and other living things. She is especially interested in the contrast between this ubiquitous improvisation and the brief bouts of planning the importance and extent of which we tend to overestimate.

Robert Rosenberger is an Associate Professor of Philosophy at the Georgia Institute of Technology in the School of Public Policy. He studies the philosophy of technology, including the development of postphenomenological theory, and the investigation of topics such as smartphone-induced driver distraction, laboratory imaging, phantom vibrations, and hostile design in public spaces. He has written the book *Callous Objects: Designs Against the Homeless*, and has edited or co-edited *Postphenomenological Investigations*, *Philosophy of Science: 5 Questions*, and *Postphenomenology and Imaging*.

Ashley Shew is an Associate Professor in Science, Technology, and Society at Virginia Tech, where she works in philosophy of technology at the intersection of disability studies, biotech ethics, animal studies, and emerging technologies. She is the author of *Animal Constructions and Technological Knowledge* (Lexington, 2017) and co-editor of three volumes in philosophy of technology: *Spaces for the Future* (with Joseph C. Pitt, Routledge, 2017), *Feedback Loops: Pragmatism About Philosophy and Technology* (with Andrew Garnar, Lexington, 2020), and *Reimagining Philosophy and Technology, Reinventing Ihde* (with Glen Miller, Springer, 2020). Supported by a US National Science Foundation CAREER Grant (#1750260), her current work centers on narratives from the disability community about technologies, especially those that serve to counter techno-ableism. She also serves as co-editor-in-chief of the journal of the Society for Philosophy and Technology, *Techné*.

Grant Tavinor is Senior Lecturer in Philosophy at Lincoln University in New Zealand. His area of research is the aesthetics of videogames and virtual reality, and his 2009 book, *The Art of Videogames*, was the first full-length work devoted to the aesthetics of games. He is editor, with Jon Robson, of *The Aesthetics of Videogames*, the first collection of philosophical papers on the topic and he has contributed essays on games to various edited book collections and to the *Journal of Aesthetics and Art Criticism*, *Philosophy and Literature*, *Contemporary Aesthetics*, and *Philosophy Compass*. His current research focuses on virtual reality media, and his next book is titled *The Aesthetics of Virtual Reality*.

Shannon Vallor is the Baillie Gifford Chair in the Ethics of Data and Artificial Intelligence and Director of the Centre for Technomoral Futures in the Edinburgh Futures Institute at the University of Edinburgh, where she is also appointed as Professor in the Department of Philosophy. Her research explores how emerging technologies reshape human moral and intellectual character, and maps the ethical challenges and opportunities posed by new uses of data and artificial intelligence. Her work includes advising academia, government, and industry on the ethical design and use of AI, and she is a former Visiting Researcher and AI Ethicist at Google. She is the author of *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (Oxford University Press, 2016), past President of the Society for Philosophy and Technology, and recipient of multiple awards for teaching, scholarship, and public engagement, including the 2015 World Technology Award in Ethics.

Peter-Paul Verbeek is Distinguished Professor of Philosophy of Technology at the University of Twente, The Netherlands, where he is also co-director of the DesignLab. He is also an Honorary Professor of Techno-Anthropology at Aalborg University, Denmark and chairperson of the UNESCO World Commission for the Ethics of Science and Technology. His books include *What Things Do: Philosophical Reflections on Technology, Agency, and Design* (Penn State University Press, 2005); *Moralizing Technology: Understanding and Designing the Ethics of Things* (University of Chicago Press, 2011); and the collection *Postphenomenological Investigations: Essays on Human-Technology Relations* (Lexington, 2015, ed. with Robert Rosenberger). He is currently one of the six Principal Investigators of a 10-year research program on the Ethics of Socially Disruptive Technologies, funded by a “Gravitation” award of the Dutch Ministry of Science.

Pieter E. Vermaas is a Professor of Philosopher of Technology at the Delft University of Technology in the Netherlands. He does research on the structure of design processes and their applications in engineering, architecture, and society at large. He co-authored the monograph *Technical Functions: On the Use and Design of Artefacts* (Springer, 2010) with Wybo Houkes. He edited the volumes *Philosophy and Design: From Engineering to Architecture* (Springer, 2008) and *Advancements in the Philosophy of Design* (Springer, 2018). Vermaas is editor-in-chief of the book series *Philosophy of Engineering and Technology* and *Design Research Foundations*. His current research extends to emerging quantum technologies and the role designing can play in the co-creation and societal exploration of the future uses of these technologies.

Kyle Whyte is George Willis Pack Professor of Environment and Sustainability at the University of Michigan. Previously, Kyle was Professor of Philosophy and Community Sustainability and Timnick Chair at Michigan State University. Kyle's research addresses moral and political issues concerning climate policy and Indigenous peoples, the ethics of cooperative relationships between Indigenous peoples and science organizations, and problems of Indigenous justice in public and academic discussions of food sovereignty, environmental justice, and the anthropocene. He is an enrolled member of the Citizen Potawatomi Nation.

D. E. Wittkower is Associate Professor of Philosophy at Old Dominion University, where he teaches on philosophy of technology, information ethics, and cybersecurity. He is editor-in-chief of the *Journal of Sociotechnical Critique*, editor or author of six books on philosophy for a general audience, and author or co-author of 46 book chapters and journal articles. He has also written for Slate, Speakeasy, and Passcode, and has recorded a dozen audiobooks that, combined, have been accessed over a million times. His current work concerns disability, trust, privacy, and the Internet of Things.

Pak-Hang Wong is formerly a Research Associate at the Research Group for Ethics in IT in the Department of Informatics, Universität Hamburg, Germany. His research explores the social, ethical, and political issues of artificial intelligence, robotics, and other emerging technologies. Wong received his doctorate in Philosophy from the University of Twente, the Netherlands in 2012 and then held academic positions in Oxford and Hong Kong. He is the co-editor of *Well-Being in Contemporary Society* (2015, Springer) and *Harmonious Technology: A Confucian Ethics of Technology* (2021, Routledge). He has published in *Philosophy & Technology, Zygon, Science and Engineering Ethics*, and other academic journals.

Aimee van Wynsberghe is Alexander von Humboldt Professor for Applied Ethics of Artificial Intelligence at the University of Bonn in Germany. Aimee is director of the Institute for Science and Ethics and co-director of the Foundation for Responsible Robotics. In each of these roles, Aimee works to uncover the ethical risks associated with emerging robotics and AI. Aimee was a member of the European Commission's High-Level Expert Group on AI and is a founding editor for the international peer-reviewed journal AI & Ethics. She is a member of the World Economic Forum's Global Futures Council on Artificial Intelligence and Humanity and author of the book Healthcare Robots: Ethics, Design, and Implementation. Aimee's current research, funded by the Alexander von Humboldt Foundation, brings attention to the sustainability of AI by studying the hidden environmental costs of developing and using AI.

CHAPTER 1

INTRODUCING THE PHILOSOPHY OF TECHNOLOGY

SHANNON VALLOR

1. An Introduction to a Late Arrival

There is something oddly anachronistic about an *Oxford Handbook of the Philosophy of Technology* first appearing in 2022. After all, technology has profoundly shaped human thought and action for as long as humans have existed. Our prehistorical forays into practices of hunting, gathering, trading, defending, and sheltering were enabled only by our ability to deploy the power of imagination in the service of *technique*: the creative manipulation and reconfiguration of objects in our physical and social environment to serve new practical ends. Studies of tool use by our primate cousins and other nonhuman animals have made it increasingly clear that the technical imagination is not limited to the human family, but a capability evolved by intelligent creatures as evolutionarily divergent from us as crows and cephalopods (Shew 2017). Nevertheless, the intimate link between technology and humanity has long been a preoccupation of historians, artists, and social and natural scientists. Why, then, should a mature *philosophy* of technology worthy of documentation have arrived so late in our history, millennia after the first self-declared 'lovers of wisdom' began writing sophisticated treatises about our aesthetic, moral, political, and epistemic capabilities?

Of course, ancient philosophers were not silent on the topic of technology. Plato and Aristotle both explored the contours of *technē* or "craft knowledge," as well as several other classifications of material production and art that encompass what today we call technology. Yet their aims in doing so were largely those of distinction, relegation, and negation; to demonstrate what separated craft knowledge from the nobler forms of the intellect: *epistēmē, phrónēsis, sophia,* and *nous.* Both Plato and Aristotle repeatedly characterize technologies and the productive arts as those forms of activity and knowledge least worthy of a philosophical mind. (Whitney 1990, 50–51). The motivation for the devaluation of the technological is usually explained in terms of the linkage between technology and the physical world, which Plato's legacy devalues in favor of the immaterial soul. Yet it is difficult not to see here another familiar pattern: the devaluation of a domain of skill and knowledge that happens to be universally reflected in the work of women and others in the domestic sphere, and brought outside the home by classes of laborers often excluded from the political elite. Indeed, while some have sought a basis in Plato for a positive relationship between the technical arts and wisdom,[1] Plato endeavors to formalize the philosophical expulsion of *technē* when he states in the *Laws* that "no citizen of our land nor any of his servants should enter the ranks of the workers whose vocation lies in the arts or crafts (846d)." The prejudice, then, is not merely of significance for metaphysics but one of profound political import.

Given that patriarchal, classist, and xenophobic hierarchies remain forcibly defended today to an extent that Plato's metaphysics will never be, one may be forgiven for concluding that the metaphysical commitment to the unchanging realm of pure ideas is not an adequate explanation for the ancient Greek prejudice against *technē*, but more plausibly interpreted as an *ex-post* rationale for the political exclusion of the artisan class. Aristotle's analysis in the *Politics* (Aristotle 2000 translation) supports this, given that he makes no mention of the exclusion being a necessary consequence of any metaphysical commitment:

> In ancient times, and among some nations, the artisan class were slaves or foreigners, and therefore the majority of them are so now. The best form of state will not admit them to citizenship; but if they are admitted, then our definition of the excellence of a citizen will not apply to every citizen, nor to every free man as such, but only to those who are freed from necessary services. The necessary people are either slaves who minister to the wants of individuals, or mechanics and laborers who are the servants of the community.
>
> (*Politics* 1278a2–12)

Aristotle's rationale here is not that those of the artisan class debase their souls with physical rather than intellectual occupations. Indeed, for both Plato and Aristotle, there are plenty of physical activities (exercise, combat) in which a proper citizen of Athens may excel. And notably, despite Aristotle's evident contempt for "slaves or foreigners," xenophobia is not the controlling rationale given, for he acknowledges that they *might* be admitted to citizenship. Rather, the truly decisive reason is their status as performers of 'necessary services': that is, persons with a willingness and refined ability to *serve and meet the needs of others.* It is the role of *serving and caring* that Aristotle finds most contemptible and beneath the status of a true citizen, and insofar as technique is historically a means of meeting a family or community's needs, it is tainted by its association with domestic care and service.

The remarkable disinterest of Greek philosophers in the study of domestic life and knowledge, and their resulting inattention to vital human realities of interdependence, service, vulnerability, and care, is a well-known defect inherited by much of the Western philosophical tradition, one noted by philosophers as distinct as Joan Tronto (1993) and Alasdair MacIntyre (1999). In addition to sustaining a philosophical prejudice against the technical arts, this defect has also, of course, served to perpetuate unjust social hierarchies and historical patterns of poor treatment of women, ethnic and racialized minorities, disabled people, and other marginalized groups. It is worth noting that more than two millennia after Aristotle's declaration of political contempt for human beings engaged in serving and caring practices, there remains an enduring link between those group identities historically targeted for social injustice—exclusion, abuse, and marginalization—and those identities most closely associated with either providing, or needing, skilled care and service, especially in the domestic realm.

While similarly unjust hierarchies are reflected in most philosophical traditions, the Greek tradition's peculiar hostility to the technical arts and relations of domestic care stands in some contrast with philosophical traditions such as Confucianism, in which a proper understanding of the structure and care obligations of family life is the building block for further modes of political and moral understanding. While technology as such is not a central theme of Confucian thought, Confucianism explicitly and consistently *valorizes* the practical embodiment of philosophical wisdom in modes of physical craft and technique, such as dress, music, dance, and calligraphy, held to have profound ethical, political, and epistemic significance. In contrast to the Greco-Roman modes of self-discipline described by Michel Foucault ([1982] 1988) as *techniques de soi* ("technologies of the self"), Confucian modes of technique do not aim to cultivate the self as a discrete soul prepared to depart the material world unencumbered, but rather as a being wholly and properly constituted by the social, political, and material dimensions of living (Gier 2001).

Yet even as craft knowledge was elevated in the social hierarchy of medieval European life by guilds of apprenticeship and mastery, technology and technique continued to be phenomena of little general interest to the modern European philosophical tradition. Early and limited exceptions include treatises such as Francis Bacon's *Novum Organum* ([1620] 2000), which explored the growing utility of scientific instruments and technique for the empirical investigation of nature. Even here, however, the power of technology to open new perspectives on nature is not identified as a topic worthy of independent investigation. Instead, technology and the technological imagination (the 'mechanical arts') are assimilated as convenient tools for two primary ends: the epistemic service of a logical method of scientific investigation and theory construction, and the practical service of increasing human comfort and 'commodious living.'

We see in Bacon's affirmation of technology's practical powers the first move away from the ancient Greek devaluation of technical intelligence as a way to serve and meet human needs. For Bacon, the power of the technical arts to serve and bring comfort to humanity is to be valorized, not scorned as Aristotle had. And yet it is worth noting that in order to valorize it, he finds it necessary to convert the valence of technique itself from servant to master. Bacon's gendered metaphors explicitly frame technology's power as a means of dominance and control over nature as a personified female other. The technical arts help science "command" nature and force "her" into the role of humanity's domestic servant by giving up the secrets she holds. Thus Bacon's philosophical valorization of the technical arts as finally worthy of noble human pursuit becomes possible only by making it compatible with a (explicitly gendered) vision of dominance and control, as noted by Carolyn Merchant:

> The material and the visual combined to produce power over nature. "By art and the hand of man," Bacon stated, nature can be "forced out of her natural state and squeezed and molded" into revealing her hidden secrets. Under the mechanical arts, he wrote, "nature betrays her secrets more fully . . . than when in enjoyment of her natural liberty." Technological discoveries "help us to think about the secrets still locked in nature's bosom." "They do not, like the old, merely exert a gentle guidance over nature's course; they have the power to conquer and subdue her, to shake her to her foundations."
>
> (Merchant 2008, 162)

Moreover, our legitimate philosophical interest in technology is for Bacon still only derivative, not primary. The philosophical value of technology is wholly contingent upon and exhausted in its contributions to the intellectual and practical achievements of the scientific enterprise; there is no further need for a *philosophy of technology*.

Along with Bacon, Galileo and Descartes' attentive interests in instrumentation had also valorized technology's scientific contributions, together marking a broader shift in sixteenth and seventeenth-century natural philosophy toward a mechanistic understanding of nature. The metaphysical rationale for the ancient Greek philosophical prejudice against the technological was evaporating; if nature *itself* is a machine, and God himself the grandest of mechanical artisans (the "Great Clockmaker"), then one could hardly sustain the Platonic characterization of the mechanical arts as *scientifically* ignoble. The sharp divide between technical artifacts and natural phenomena blurs; even if the hierarchy is retained by the qualitative distinction between divine and human artifice.

Yet most philosophers continued to be remarkably silent on the topic of technology even as the Scientific and Industrial Revolutions rebuilt the world around them. While many remained committed to a dualist or idealistic metaphysics that would permit the retention of the ancient prejudice against the technical arts, it is noteworthy that among the few late modern philosophers to pick up the theme in ways that extended beyond a theory of scientific instrumentation was Karl Marx. Marx's treatment of technology is broadly recognized for its nuance and ambiguity, insofar as he recognized its polyvalent potential to both alienate and liberate human beings. Yet despite his extensive discussion and historical study of technology as a force in human affairs, Marx's thought is also not yet a *philosophy of technology*, for it remains constrained by the focus of his broader project. Technology for Marx is the *machine*, a phenomenon of vital significance in the context of modern labor systems, but the roles it might play outside of the forces of production and capital are largely unexplored (Wark 2019). Yet in Marx we do find the first suggestions of a more encompassing philosophy of technology:

> Technology reveals the active relation of man to nature, the direct process of the production of his life, and thereby it also lays bare the process of the production of the social relations of his life, and of the mental conceptions that flow from those relations.
>
> (Marx [1867] 1976, 493)

Here Marx points to the fundamental role that technology plays not only in the material production of life, but also in social, cultural, and cognitive production. Shortly thereafter, Ernst Kapp's *Grundlinien einer Philosophie der Technik* (1877) would establish the first mature work in the philosophy of technology in Europe, a tradition that retains a rich and enduring legacy (Dessauer 1927; Ortega y Gasset 1939; Simondon [1958] 2016; Anders 1956). But these works were not widely read or translated beyond the continent, and it would take nearly a century for technology to draw the focal attention of the English-speaking philosophical community.

With the exception of John Dewey's diffuse ruminations on the topic (Hickman 1990) and the work of Lewis Mumford (1934), Anglophone philosophical interest in technology remained subdued and ephemeral in the early twentieth century. For better or worse, a robust English-language philosophy of technology awaited the postwar development of the themes established in Martin Heidegger's brief but highly influential treatment of the topic in "The Question Concerning Technology [*Die Frage nach der Technik*]" ([1954] 1977). Among readers already engaging with continental phenomenology while culturally enveloped by a postwar expansion of scientific, military, and industrial powers, Heidegger's concept of modern technology as a dangerous "Enframing" (*Gestell*) of all reality as manipulable, calculable, and exploitable resources or "standing reserve" (*Bestand*) struck a deep chord. The scholarship that began to grow from this conception of the essence of technology as a world-historical force soon dovetailed and merged in intricate ways with a parallel line of analysis influenced by critical theory of the Frankfurt School, which picked up Marx's thread and extended it in new directions (Marcuse 1964), while incorporating insights from sociological and historical perspectives on technology (Mumford 1934, Ellul [1954] 1964) that now seemed all the more philosophically fecund.

The result was the long-delayed coalescence of English-language philosophy of technology as a field; that is, a community of Anglophone philosophical readers who shared (at least initially) a common conceptual frame, lexicon, canonical literature, and set of problematic questions about technology to pursue. Founded in 1976, the Society for Philosophy and Technology gave the first formal unity to an international community of academics for whom technology was not merely a derivative interest in service to philosophy of science or political philosophy, but was a subject worthy of philosophical understanding in its own right. In 1995, the society established its own journal, *Techné: Research in Philosophy and Technology.* Yet the self-destructive and provincial tendencies engendered by the growing ‘analytic-continental divide’ in postwar academic philosophy meant that late twentieth century philosophy of technology, marked as it was by its close affiliations with continental phenomenology and critical theory, would for a time suffer from the same sort of intellectual self-confinement that has afflicted many analytic philosophical communities in parallel. The philosophy of technology might have become quite insular and sterile, in fact, were it not for three further contemporary developments.

The first was the growing engagement of philosophy of technology with the interdisciplinary field of STS (Science and Technology Studies), enabling new methodologies, literatures, and conceptual framings to be injected into both discourses. Just as analytic philosophy of science began to stretch beyond its narrow logical preoccupations to respond to sociological, historical, and political questions posed by the STS literature, continental philosophers of technology were compelled to do the same, resulting in insights newly enriched by the social and historical perspectives of scholars such as Bruno Latour and Steve Woolgar. The second shift was a growing consensus that, following successive revelations of the depth and apparent philosophical sincerity of Heidegger’s anti-Semitism (Mitchell and Trawny 2017), the field needed to extricate itself from reliance upon his philosophical framing. The third and most recent shift has been the evident and growing need for philosophy of technology as a practical tool of analysis, one that can aid technologists, engineers, policymakers, and technology users in better understanding and more wisely shaping the core technological dimensions of social life. Thus by the start of the twenty-first century, philosophy of technology had begun to move away from essentialist inquiries about technology as a dimension of human nature and overarching force in history, and toward questions about the meaning and empirical significance of specific, contemporary technoscientific developments in fields ranging from biotechnology and nanotechnology to computing, cognitive science, and robotics—a phenomenon that came to be known as the ‘empirical turn’ (Achterhuis 2001).

The chapters that follow, as described more fully in the next two sections, largely trace the living problems of the field in the wake of the empirical turn. These are multiple and rich. But it must be noted that the field has yet to fully confront the contingencies and omissions of its peculiar history. With rare exceptions, the ancient Greek philosophical project and the contemporary European and Anglo-American philosophies descended from it remain marred by a deep and latent intellectual bias against the technological that has allowed our understanding of this dimension of our nature to remain occluded and partial, even as it became the most distinctive and transformative feature of modern and contemporary life. Moreover, insofar as the technological has been confronted as a philosophical question, the question remains largely framed only in light of a moribund hierarchy in which the productive force of technology in games of economic, military, and scientific domination still claims the highest rung. The many roles that technologies and the technical imagination play in the broader aesthetic, moral, psychological, domestic, and cultural dimensions of human life remain profoundly undertheorized in academic philosophy, especially when compared with far richer and more expansive bodies of work in sociology, anthropology, and history of technology and technological practices.

The consequences of these omissions go beyond intellectual inadequacy. They have inhibited the emergence of new philosophical visions of our human relationship to technology: such visions as might see a way forward to new, more socially and environmentally sustainable futures than those toward which we remain relentlessly anchored by our clinging to outmoded, impoverished hierarchies of value and human worth. In the kinds of technomoral futures we might have envisioned instead, technologies would be conceivable as more than deterministic engines of economic production, exploitation, and political domination whose harms and risks can at best only be mitigated and contained. In a different world, built with more expansive and inclusive philosophical imaginations, technologies and technique could have been seen—and might still be seen—as expressions of human freedom in solidarity with others, even as new avenues for the materialization of *love*. Technologies can be, and often have been, engines not merely of war and wealth, but also of creative play, artistic expression, social care, service, and comfort to others. Many of the chapters herein explore the renewal and expansion of such possibilities, and in them, I believe, lie our best philosophical hopes for finally, belatedly, becoming wise about technology.

2. The Structure of the Volume

A volume such as this admits of many possible principles of organization, and no one model is obviously superior. The structure of this particular Handbook was chosen in order to reinforce the importance of a philosophy of technology not merely as a narrow subspecialty, but as a cross-cutting inquiry that does and must inform nearly every other established domain of philosophy, while drawing from and contributing to closely related fields including science and technology studies (STS), anthropology, history and sociology of technology, and media and information studies.

The methodologies, vocabularies, and literatures employed in the sections and chapters herein frequently cut across the continental and analytic, conceptual and empirical, theoretical and practical divides. Contributors range from early founders and established leaders of the English-speaking research community to emerging scholars opening entirely new horizons of research. The chapters, themes, and perspectives in this volume thus represent a rich, diverse, and expanding subject area of vital and growing importance to scholars, policymakers, and practitioners worldwide—while the field itself still suffers, as do virtually all Anglophone philosophical communities, from acute failures of representation and inclusion; a theme that is itself taken up by a number of authors.

Each of the seven sections of this volume reflect a familiar and well-established thematic cluster of philosophical research, so that those new to the philosophy of technology, who might open this volume wondering where technology “fits” in the traditional philosophical landscape, have their answer: *everywhere.* There is no complete epistemological theory, philosophy of mind, philosophy of science, or history of philosophy itself that does not account for the role of technologies and technique in the production and justification of human knowledge, thought, experience, and wisdom. There is no adequate political theory of justice that does not incorporate an understanding of technology as a source and expression of both human power and freedom. There is no comprehensive ethics that does not address technology's role in shaping and materializing human values in action, and in constructing and negotiating multiple visions of the good life. In addition to these intersections, the sections that follow reveal the philosophy of technology as a vital domain of understanding for philosophical metaphysics, aesthetics, and environmental philosophy.

Yet all choices carry a cost, and this particular organizational framing leaves unfulfilled at least one other possibility: that a mature philosophy of technology might show us how to reconstruct our familiar philosophical taxonomy and traditional lexicon in new and more fruitful ways. I leave it open to this volume's readers, and to new generations of philosophers, to consider whether or how this might be done.

3. Overview of Sections and Chapters

Section One: Histories and Methodologies in the Philosophy of Technology

This section of the volume explores the *histories and methodologies* that have shaped the development of philosophy of technology as a field and that condition its present prospects and possible future trajectories. The historical orientation of these chapters is not one of passive survey; each one is carried out in the mode of active and vital contestation of the boundaries, legitimate aims, and conceptual tools of the field. This section thus forcefully engages multiple controversies that mark the field's temporal and conceptual arcs, including the relationship between continental and analytic philosophy of technology; the divergence of the 'empirical turn' from prior essentialist analyses; the proper role of critique in technology ethics; and the means by which the philosophical legitimacy and practical contributions of the field might best be secured.

Chapter 2, by Carl Mitcham, revisits three twentieth century European thinkers who have profoundly shaped our understanding of technology: Alan Turing, Jacques Ellul, and Martin Heidegger, in order to ask in which of these "classic" works, now often read as merely historical legacies to be surpassed, we might discover still-living issues for the field. In Chapter 3, Peter-Paul Verbeek asks a similar question from the reverse of this temporal angle, focusing on the early twenty-first century 'empirical turn' in philosophy of technology that sought to leave its twentieth century essentialist legacy behind. Verbeek asks how further "turns" might build upon the empirical transition from a philosophy *of* technology, to philosophy *from* and *for* technology. Chapter 4 engages with the impact of the 'continental/analytic divide' on philosophy of technology's past and present. In it, Maarten Franssen argues that this divide has contributed to the delayed maturity of the field and its failure to consolidate into a coherent whole, proposing that a further tripartite division of the field might be necessary to secure its future. Finally, Chapter 5 offers a provocation from Don Howard, who argues that the field remains inhibited by an immature and reactionary techno-pessimism inherited from its twentieth century roots, one that must be rejected in favor of philosophical collaboration with scientists, technologists, and policymakers who seek to embrace technology's amelioratory possibilities for human flourishing.

Section Two: Technology and Epistemology

Here we explore the links between technology and epistemology, broadly construed, beginning with A. S. Aurora Hoel's analysis in Chapter 6 of the epistemic role of scientific instruments as *adaptive mediators* of knowledge. She draws upon Gilbert Simondon's model of ecological relationality to argue for the prospect of new epistemologies centered on technological mediation that help us transcend the sterile, deadlocked framing of the 'science wars.' In Chapter 7, Wybo Houkes and Anthonie W. M. Meijers articulate a philosophy of engineering knowledge that integrates recent work in the philosophy of science to enable a new understanding of the epistemic activities, rules, and values that govern design.

Chapter 8 explores the epistemic import of Beth Preston's analyses of the *technical functions* of artifacts; in it, Preston draws upon the 'continuum problem' in classifying technical functions to defend a view of scientific classification as an epistemic rather than ontological project. Section Two concludes with Sage Cammers-Goodwin's use in Chapter 9 of standpoint epistemology and 'hostile design' literature to critique the tacit and exclusionary assumptions present in contemporary 'smart city' discourse and initiatives, which systematically devalue or omit the specialized knowledge possessed by city dwellers, especially marginalized citizens whose needs and understandings of city life are systematically ignored.

Section Three: Technology, Power, and Politics

This section explores the intersections between technology, power, freedom, justice, and identity. It begins with Chapter 10's meta-analysis by Adam Briggle of the politics of philosophy of technology as a field, offering both a defense and a critique of the technique(s) of philosophy of technology itself, and suggestions for future reform. In Chapter 11, Alison Adam deploys a postcolonial critique of traditional accounts of how scientific and technological knowledge emerge and travel, drawing upon historical and sociological analysis of the development and dissemination of technologies of identification in colonial India. Chapter 12 provides a Rawlsian analysis of information technologies' impact upon social justice and identity; in it, Anna Lauren Hoffmann develops a notion of the "sociotechnical bases of self-respect" to better account for how information technologies condition opportunities for dignity and justice in liberal democratic societies.

Chapter 13 offers John Danaher's account of the rise of 'algocracies'—modes of algorithmic governance that challenge utopian hopes about the liberatory potential of digital technologies. Danaher argues for a broader conception of both freedom and algocracy that can clarify both the emancipatory and oppressive possibilities of algorithmic governance. Finally, in Chapter 14, Anna Gotlib argues for a refocusing of political philosophy of technology upon the identity-constituting effects of technological innovations, especially in the domain of biotechnology, where the identities of the vulnerable and oppressed stand in particular danger.

Section Four: Technology, Metaphysics, and Language

This section offers diverse accounts of the role technologies play in constituting ourselves and our realities, beginning with Ciano Aydin's analysis in Chapter 15 of the 'technological uncanny' evoked in the well-known phenomenon of the 'uncanny valley' produced by humanoid robots. Aydin draws upon empirical studies of the phenomenon along with existential and psychological perspectives from Freud, Lacan, and Nancy to develop a new explanation of the technological uncanny as pointing, not to an emotional or evolutionary affront by the humanoid, but to an existential-metaphysical gap in our own self-understanding. In Chapter 16, Massimo Durante turns the metaphysical lens upon the virtual domain, drawing upon Floridi's philosophy of information to reveal how our traditional ontologies and metaphysical frameworks are (or are not) challenged by the digital enabling of new virtual and hybrid or 'mixed' realities.

Chapter 17 turns to exploring the undertheorized intersections between technology and language. Mark Coeckelbergh analyzes what philosophers who study the former (especially in the post-phenomenological tradition) can learn from those who study the latter, and how bridges from the work of Searle, Ricoeur, and Wittgenstein on language can help us better understand technology and the material *and* linguistic mediation of human-technology-world relations. In Chapter 18, D. E. Wittkower develops a post-phenomenological account of artificial virtual assistants and 'bots' such as Alexa. Wittkower draws upon Nagel and Dennett's ideas about mind and intentionality to reveal how the functional design of agents like Alexa demand that users consistently attribute intentional states to machines that they cannot possess; the result is a growing practice of holding and acting upon fictitious theories of mind about such agents, with affordances and consequences we have yet to fully understand. Finally, Robert Rosenberger's contribution in Chapter 19 revisits Ihde's postphenomenological account of technological ‘multistability’ to open radical new questions about the philosophical project of seeking to return to the ‘things themselves’ in an era that generates ever-new ontological relations with technologies.

Section Five: Technology, Aesthetics, and Design

This section explores the role of technology in our experience of the built world. In Chapter 20, Philip Brey provides an account of engineering design as a human practice, and in its normative dimension as a practice that embeds moral, social, and political values and choices—with implications both for the notion of “good design” and the “good society.” Chapter 21 turns to the aesthetic dimensions of design and experience enabled by new virtual reality media; in it, Grant Tavinor explores their novelties and their continuities with pre-digital techniques in artistic perspectival depiction.

Chapter 22 returns to the subject of engineering design, this time in the context of the evaluation, validation, and management of design outcomes and methods. Pieter Vermaas draws upon the varied literature in design thinking and methods, using a case study in urban design of an entertainment district to demonstrate the complex challenges of design evaluation and management and the need for philosophers of technology to attend more carefully to them. The section concludes with Sanna Lehtinen’s analysis in Chapter 23 of the intersection of philosophy of technology with urban aesthetics; Lehtinen argues that a more robust philosophical understanding of how new technologies condition and transform our aesthetic experience and opportunities in urban environments can yield important dividends for the future of city life.

Section Six: Technology, Health, and the Environment

This section focuses on the ways in which recent technological narratives, choices, and affordances have shaped and continue to shape our relationships to living and natural systems in our own bodies and ecosystems. In Chapter 24, Julia D. Gibson and Kyle Powys Whyte explore the narrative dimensions of environmental futurism in both philosophy of technology and science fiction, and how these each project underexamined and limited concerns, assumptions, and values into the lives of subsequent generations increasingly threatened by the climate crisis. Revealing the imaginative constraints of industrialization, capitalism, and colonialism on our powers of vision, Gibson and Whyte consider how more inclusive futurist narratives in philosophy and fiction might do better at heeding the increasingly urgent call for global climate justice.

Chapter 25 turns to the domain of agricultural biotechnology to examine the policy and political challenges of emerging technology governance in domains where deep and enduring value conflicts repeatedly block compromise and cooperation. Samantha Noll employs a case study of North American debate over genetically modified organisms

(GMOs) to develop a practical framework for "unfreezing" value conflicts that obstruct fruitful consensus in environmental and biotechnology policy.

In Chapter 26, Ashley Shew employs a synthesis of philosophy of technology and disability studies to critique naïve and exclusionary narratives in futurist and transhumanist discourse about the utopian prospects of 'cyborg' transformations of human bodies. She argues that only by reprioritizing the long-neglected lived experiences and expertise of existing cyborgs, that is, disabled people with already-technologized 'bodyminds,' can we guide ourselves wisely into our futures. Chapter 27 concludes the section by exploring the vast range of ethical quandaries posed by the emerging prospect of widely expanded human activity in outer space. From commercial space tourism to space militarization to terraforming and space colonization, Keith Abney reveals both the vast new frontiers of space ethics and the troubling complexities that arise when new technology enables previously Earthbound ethical debates to be transported into non-terrestrial environments.

Section Seven: Technology and the Good Life

This section concludes the volume with what can only be an incomplete sample of a theme that has arguably occupied the vast majority of philosophers of technology in one way or another. Indeed, the relationship between technology and ethics—whether in the context of justice, freedom, identity, sustainability, health, responsible design and engineering, or sound public policy—is already woven throughout many other chapters in the volume, as it must be. For if anything true can be said of technology as a whole, it is that it expresses and materializes human values and needs at every turn, even when it is harmful or misused. Thus ethics is never far away from a philosophy of technology.

This final section, then, does not survey or confine this domain. Indeed, the task would be impossible, as entire Handbooks have been dedicated to the research within single subareas of technology ethics, such as the *Oxford Handbook of Ethics of AI* (Dubber et al. 2020) and the *Oxford Handbook of Digital Ethics* (Véliz, forthcoming). Instead, Section Seven reflects a particular lens we may use to look at the ethics of technology as a whole. Its focal point is not the level of individual right action in the present technosocial milieu, but the level of human flourishing as an unrealized set of technosocial possibilities that lie in our future. The closing section of this volume is devoted to possible futures in which human moral character, virtue, care, and community are recognized as necessary conditions of living well with technology.

The section begins with Barbro Fröding's analysis in Chapter 28 of the prospects of emerging technology-enabled cognitive enhancement for human flourishing. Drawing upon examples of cognitive enhancement by means of computer training, transcranial direct stimulation, neuro/biofeedback, and brain-computer interfaces, Fröding proposes virtue ethics as a motivating and constraining condition of the ethical development and use of cognitive enhancement technologies. In Chapter 29, Charles Ess turns to the existential legacy of the Epic of Gilgamesh to draw out an philosophical anthropology in which the emancipatory conditions of flourishing with technologies are revealed through our nature as embodied, relational beings. Synthesizing perspectives from Enlightenment, Romantic, feminist, and virtue ethical accounts of human flourishing with technology, Ess argues for the cultivation of new courage to "hack" and reshape more liberated and liberating relationships to technology.

In Chapter 30, Pak-Hang Wong reflects on recent attempts to introduce Confucian value perspectives into the ethics and philosophy of technology. Wong argues that these attempts have yet to adequately incorporate the central role of ritual (*Li*) in Confucian thought, and demonstrates how renewed attention to the aesthetic, formative, and communicative functions of *Li* can answer the question of why Confucianism matters to technology ethics. In Chapter 31, Aimee van Wynsberghe uses an analysis of the ethics of humanitarian uses of robots to show how feminist care ethics can supply a more robust and satisfactory normative orientation to the philosophy and ethics of emerging technologies. Finally, Chapter 32 concludes the volume with Deborah G. Johnson's analysis of how the conditions of the good life with technology have been consistently framed in terms of the weighing of emerging technology's promises and perils. By deploying the sociotechnical systems perspective, Johnson explores how such framings are challenged by the inherent uncertainties that continue to limit our vision of technological futures.

4. A Conclusion, an Indictment, and a Call

The philosophy of technology had a late start, and it still carries with it the unresolved tensions, ambiguities, and anxieties about technology's relationship to humanity and the good life that have nagged us since Plato first set *technē* aside from wisdom. The accelerating technosocial transformations of the modern and contemporary era, and their yet unreconciled consequences for human political life and the sustainability of the planet, have only amplified philosophical anxieties and unease about technology.

And yet this is precisely why philosophers—in far more expansive and inclusive numbers—are called, now more than ever, to attend to technology. The price of ignoring *technē* as beneath philosophical interest, of laboring under the illusion that technology consists of "mere tools" of ephemeral consequence rather than fundamental dimensions of our existence, has already been far too high. The most acute cost by far has been philosophy's failure to equip the human family with the habits of mind and action needed to competently manage, much less prevent, the existential threats to our polities and our planet currently posed by our long-disordered technological practices: practices that have for millennia been allowed, by active neglect as much as by circumstance, to develop in a manner that is profoundly *mindless*.

This is not an indictment of technology—but it is a stain on philosophy, as a human cultural practice devoted to the *prevention* of mindless living. And there are no guarantees to be had that it is not too late. As philosopher and theologian Hans Jonas observed decades ago:

> With the apocalyptic pregnancy of our actions, that very knowledge which we lack has become more urgently needed than at any other stage in the adventure of mankind. Alas, urgency is no promise of success (1973, 52).

There are also no guarantees that the learned, privileged, and often inward-looking habits of present-day academic philosophy are still the habits we need. Is the kind of wisdom Jonas calls for, the kind that with courage and humility heeds the call not only to knowledge but to *responsibility*, still living and vital within the philosophical community? If not, where may it be found?

> This raises to an ultimate pitch the old question of the power of the wise, or the force of ideas not allied to self-interest, in the body politic. What *force* shall represent the future in the present?
>
> (Jonas 1973, 51)

Philosophy still has a role to play in representing the future, though its claims to be able to do so justly and wisely are profoundly diminished in legitimacy by the gatekeeping that still prevents our discipline's academic membership from representing the human family in any meaningful sense. Academic philosophy would need to constitute a very different community of practice before it could be judged to represent the body politic, much less the "force of ideas not allied to self-interest" within it. And if the "power of the wise" were to at last begin to shape our futures with technology, it would not be philosophers alone, but creators, leaders, and community voices of every stripe who would embody it. Most important, it would have to be guided by the wisdom of those who Plato and Aristotle long ago expelled from both philosophy *and* power—persons who embody the art of technique in the *necessary service and care* of others.

Thus humility and respect, virtues with which philosophers characteristically struggle (to put it mildly), must become far more constant companions of our pursuit of wisdom. Yet, in the absence of adequate philosophical engagement with technology, its rigorous academic study will continue to be left to empirical sciences that can only tell us what human artifacts and techniques *are* and *have been,* how they *are* and *have been* embedded in our lives—but not what our lives with technology *might* be, or how they *should* be freed to become, for us and for future generations. The contributions in this volume reflect an enduring and growing communal effort to respond to this urgent call, one that I hope many new others will join in the living spirit of philosophical wisdom.

Note

[1] See Roochnik (1998) on attempts to reconcile the technical arts and moral wisdom in Plato, which he concludes are unsuccessful.

References

Achterhuis, Hans (ed.) 2001. *American Philosophy of Technology: The Empirical Turn*. Bloomington, IN: Indiana University Press.

Anders, Günther. 1956. *Die Antiquiertheit des Menschen*. Munich: C.H. Beck.

Aristotle. 2000. *Politics*. Translated by Benjamin Jowett. Mineola, NY: Dover.

Bacon, Francis. [1620] 2000. *The New Organon*. Cambridge: Cambridge University Press.

Dessauer, Friedrich. 1927. *Philosophie der Technik: Das Problem der Realisierung*. Bonn: Friedrich Cohen.

Dubber, Markus D., Frank Pasquale and Sunit Das. 2020. *The Oxford Handbook of Ethics of AI*. New York: Oxford University Press.

Ellul, Jacques. [1954] 1964. *The Technological Society*. New York: Vintage Books.

Foucault, Michel. [1982] 1988. "Technologies of the Self." In *Technologies of the Self: A Seminar with Michel Foucault*. Edited by Luther H. Martin, Huck Gutman, and Patrick H. Hutton, 16–49. Amherst, MA: Univ. of Massachusetts Press.

Gier, Nicholas F. 2001. "The Dancing *Ru:* A Confucian Aesthetics of Virtue." *Philosophy East and West* 51(2): 280–305.

Heidegger, Martin. [1954] 1977. "The Question Concerning Technology. In *The Question Concerning Technology and Other Essays*, 3–35. Translated by William Lovitt. New York: Harper & Row.

Hickman, Larry. 1990. *John Dewey's Pragmatic Technology*. Bloomington, IN: Indiana University Press.

Jonas, Hans. 1973. "Technology and Responsibility: Reflections on the New Tasks of Ethics." *Social Research* 40(1): 31–54.

MacIntyre, Alasdair. 1999. *Dependent Rational Animals: Why Human Beings Need the Virtues*. London: Bloomsbury.

Marcuse, Herbert. 1964. *One Dimensional Man: Studies in the Ideology of Advanced Industrial Society*. Boston: Beacon Press.

Marx, Karl. [1867] 1976. *Capital: A Critique of Political Economy. Vol. I*. Harmondsworth: Penguin.

Merchant, Carolyn. 2008. "Secrets of Nature: The Bacon Debates Revisited." *Journal of the History of Ideas* 69(1): 147–162.

Mitchell, Andrew J. and Peter Trawny. 2017. *Heidegger's Black Notebooks: Responses to Anti-Semitism*. New York: Columbia University Press.

Mumford, Lewis. 1934.*Technics and Civilization*. New York: Harcourt.

Ortega y Gasset, José. [1939] 2012. *Filosofia Contemporánea: "Meditacion de la technica."* Valencia: Boreal Libros.

Roochnik, David. 1998. *Art and Wisdom: Plato's Understanding of Technē*. University Park, PA: Penn State University Press.

Shew, Ashley. 2017. *Animal Constructions and Technological Knowledge*. Lanham, MD: Lexington Books.

Simondon, Gilbert. [1958] 2016. *On the Mode of Existence of Technical Objects*. Minneapolis: Univocal Publishing.

Tronto, Joan. 1993. *Moral Boundaries: A Political Argument for an Ethic of Care*. London: Routledge.

Véliz, Carissa, ed. (forthcoming). *The Oxford Handbook of Digital Ethics*. Oxford: Oxford University Press.

Wark, McKenzie. 2019. "Technology." In *The Bloomsbury Companion to Marx*, edited by Jeff Diamanti, Andrew Pendakis, and Imre Szeman. 621–628. London: Bloomsbury.

Whitney, Elspeth. 1990. *Paradise Restored: The Mechanical Arts from Antiquity Through the Thirteenth Century*. Philadelphia: American Philosophical Society.

PART I

HISTORIES AND METHODOLOGIES IN THE PHILOSOPHY OF TECHNOLOGY

CHAPTER 2

WHAT IS LIVING AND WHAT IS DEAD IN CLASSIC EUROPEAN PHILOSOPHY OF TECHNOLOGY?

CARL MITCHAM

Recently a Dutch colleague and I were commiserating about our overflowing bookshelves and how we couldn't keep up with everything being published about philosophical issues related to technology. We wondered whether it might not be just as good to go back and re-read the books already in our libraries. We agreed that in post-World War II Europe, technology as a philosophical problem began to precipitate out in different ways in different linguistic communities, from where it drifted into more general philosophical discourse, even if the term "philosophy of technology" didn't arrive until later. But there we left it.

A few months later I was drafting the syllabus for a graduate seminar introducing students at Renmin University of China to Western philosophy of technology. Recalling the earlier discussion, I decided to take seriously the hypothesis that some works resting on my office shelves might have more salience than they were receiving today amidst the publishing frenzy that has engulfed the academic world, especially as transformed by institutional demands, digital dissemination, and social media. As someone who came of philosophical age during the 1960s questioning of American technopolitical hegemony, I had initially relied for guidance on a set of thinkers who have since become somewhat marginalized. Under the banners of "postphenomenology" (Ihde 1993) and an "empirical turn" (Kroes and Meijers 2000; Achterhuis 2001), along with analytic pragmatism, in many English-speaking quarters the philosophy of technology has become narrowed down to the analysis of particular cases having to do with particular technologies. My own work had become centered in engineering ethics. So I decided as a pedagogical experiment to conduct a graduate seminar dedicated to what may be called "classic" (if not exactly canonical) European texts in philosophy of technology, inquiring to what extent they continued to present living issues. What follows reflects in part responses from the Chinese students in that seminar during the spring semesters of 2018 and 2019 as they joined me in looking back at a peculiarly fertile period in the emergence of philosophical engagement with technology.

1. What Is Classic European Philosophy of Technology?

"Classic" is a contestable term. Here classic European philosophy of technology is anchored in the recognition of modern engineering and technology as a historically unique, science-associated form of designing, producing, and using artifacts that began with the Industrial Revolution and has since progressively transformed itself and the world. Efforts to think critically rather than promotionally about this mutation in the means of production and use can be traced back to Jean-Jacques Rousseau (1712–1778), Jeremy Bentham (1748–1832), Robert Owen (1771–1858), and Karl Marx (1818–1883) and led eventually to a privileging of "technology" as a socio-cultural force. The conceptual focus blossomed in the 1950s through the 1970s primarily in the United Kingdom, Germany, France, and the United States—only to undergo critical trimming if not deflation. Hence the question: To what extent is, are, or ought work at the root of this initial flowering continue to be studied? Is this philosophy past what post-classical philosopher of technology Don Ihde (2018) calls its shelf life?

Consider three key texts from England, France, and Germany: Alan Turing's "Computing Machinery and Intelligence" (1950), Jacques Ellul's *La Technique ou l'Enjeu du siècle* (1954), and Martin Heidegger's "Die Frage nach der Technik" (1954). Each represents a different approach not just to technology but to philosophy, and can reasonably be considered classic by virtue of referential persistence. It was a pivotal time, from which we look forward and backward. Turing did conceptual analytic work on computing and information technology. Ellul developed a theory of society transformed by technology. Heidegger advanced a phenomenological reflection on *Technik* leading to ontological claims. They thus initiated traditions of analytic, social-political, and metaphysical philosophy of technology that continue to cast shadows into the present. But to what degree do the particulars of that originary work still nurture philosophical engagement with technology? Or is their work better simply referenced then left behind?

2. Alan Turing and Artificial Intelligence

Although he died young (age 42, perhaps by suicide) and did not publish much, in popular culture Turing is undoubtedly the most living of the three. As a minor contributor to the Allied victory in World War II who was nevertheless punished for his homosexuality, he has been the subject of a biography (*Alan Turing: The Enigma*, 1983), a play (*Breaking the Code*, 1986), a TV film (*Codebreaker*, 2011), an Oscar-winning movie (*The Imitation Game*, 2014), an opera (*The Life and Deaths of Alan Turing*, 2014), and numerous studies. Philosophically, Turing was a member of the highly influential community of scholars centered around Ludwig Wittgenstein, with whom he argued about issues in the ontology of mathematics. His formulation of the imitation game as a substitute for the question "Can machines think?" remains a standard trope in analytic philosophy of artificial intelligence (AI).

The paper that develops the imitation game is a modest 27 pages divided into seven sections (Turing 1950). The first three simply describe the game, defend its operationalizing of a vague or ambiguous question, and define its boundaries. Read with hindsight, the fact that the game was initially outlined as the interrogation of a man (A) and a woman (B) by an interrogator (C), who was tasked with determining which is which, when A and B are free to dissemble, cannot help but be interpreted in psychological terms. In Turing's adaptation, B is replaced by a computer and C is tasked with determining through textual interrogation alone which is human. (Through text alone, could C ever average better than 50% on the original game?) Sections four and five of Turing's paper add specifications with technical details about digital computers. Section six, the longest section by far (close to half the paper), considers a series of nine objections. A final and second longest section considers the possibility of learning machines.

The nine objections provide a convenient framework for assessing philosophical viabilities in this classic text. The first two objections didn't have much life in them even for Turing. A "theological objection" was that God gave humans but not computers a thinking, immortal soul. A "heads in the sand" objection had computers as just too scary to think about.

The third "mathematical objection" suggested that Kurt Gödel's incompleteness theorem might make the algorithmic imitation of human thinking impossible. According to Gödel, no formal axiomatic system strong enough to model basic arithmetic is also able to prove all arithmetic truths. Philosopher J. R. Lucas (1961) and physicist Roger Penrose (1989) developed versions of the argument to deny the possibility of "strong AI" (a computer with truly human cognitive abilities). The issue plays a role in neuroscientist Douglas Hofstadter's popular science book on *Gödel, Escher, Bach* (1979). In this narrow form it is nevertheless today a challenge on life support.

A fourth "argument from consciousness" (along with a fifth "argument from various disabilities" which contains "disguised forms of the argument from consciousness") remains very much a living issue in the philosophy of mind. Turing here anticipated debates on the possibilities of machine consciousness that remain basic to the philosophy of computers spanning the thought of Hubert Dreyfus and John Searle to David Chalmers and Daniel Dennett.

The next two arguments also remain living issues: "Lady Lovelace's objection" is that computers are just dumb machines that can only do what they are programmed to do. An "argument from continuity of the nervous system" asks whether the output of the human brain, which is not a discrete state machine (with clear on-off registers), can be imitated by a computer, which is a discrete state machine. These issues too continue as topics in philosophical debates, both technical and popular, about AI, both strong and weak.

An "argument from the informality of behavior" turns on a distinction between laws for the operation of a computer and rules for human behavior, rules which may or may not be followed. A strong version of this asserts that humans are free and computers are not. The issue thus becomes one between freedom and determinism, another continuing topic related to the technological modeling and possible control of human behavior.

The last "argument from extra-sensory perception" is the strangest of the nine but, remarkably, the one Turing maintained was "quite a strong one" (Turing 1950, 453), which makes it even more strange. Turing asserts that at least for one form of extra-sensory perception, "the statistical evidence . . . is overwhelming" (453). This objection is clearly a dead issue, except among new age enthusiasts.

Following review of these nine objections, Turing writes that he has spent so much time considering objections because he "has no very convincing arguments of a positive nature to support [his] views" that "in about fifty years' time [computers will] play the imitation game so well" that an interrogator will not be able to correctly distinguish between computers and humans more than 70 percent of the time after five minutes of questioning (442). Since he thinks the main barrier to reaching this goal will not be hardware development but software, he focused the rest of his article on the theory of "learning machines"—a topic that remains vitally engaging.

Post-Turing computer research underwent inflationary development (including coining of the term "AI") but then during the 1970s experienced a period of retrenchment and reduced funding (the "AI winter") followed by renewed optimisms. Philosophical interests have tended, with some lag time, to track that and subsequent cycles. In the process, information machines have become increasingly recognized as topics of significant epistemological and ontological interest, from discussions of human-computer symbiotic cognition to cyborg transhumanism. Turing was clearly a pioneer in this area so that at least historical appreciation of related contemporary discourse can be enriched by revisiting his work.

Moreover, the *élan vital* of Turing's analytic practice is such that it has migrated from computers and information technology into a host of other fields: nuclear weapons and power, biomedicine, engineering professionalism, environmental engineering, communications and media, biotechnology, genetic engineering, nano-engineering, and more. In each area specific issues are raised, subjected to conceptual refinement, and assessed by considering the strengths and weaknesses of arguments related to questions, most commonly, of knowledge or ethics.

In Turing, however, the topic of ethics is conspicuously absent. Turing did not consider any moral objections to making computers think or some of the ethical complications that might follow the successful design and construction of imitation game–winning computers, much less the emergence of automatons and intelligent robots. Here another engineer from the same period was the leader: Norbert Wiener

(1894–1964). Wiener's *The Human Use of Human Beings*, published the same year as Turing's paper, initiated another whole dimension of philosophical reflection on AI. Research in computer ethics has in fact been the leading area of philosophical growth, especially in conjunction with the increased infusion of "smart devices" into an internet of things and the ever more expansive quotidian processing of big data. In these areas Turing is less relevant, except insofar as his personal contributions to the earliest form of computer surveillance warfare can be looked back on as an anticipation of threats to come in civilian as well as military affairs.

3. Ellul and the Technological Society in France and in America

Jacques Ellul was heir to a tradition of thinking found in such classical sociologists as his compatriot Emile Durkheim (1858–1917). Like Durkheim, Ellul was born in Bordeaux, where, unlike Durkheim, he remained during a long professional and extremely productive life. However, in part because he did not move to Paris, he remained marginal among French intellectual elites, despite authoring a multi-volume textbook, *Histoire des institutions,* used in the law curriculum nationwide. His insertion into philosophy of technology occurred in the context of a historically contingent popularity in the United States among one of two mostly non-overlapping groups: evangelical Christians and left-leaning political activists (see Mitcham 1993).

This bimodal attention reflects what Ellul himself described as the dialectical character of his thinking. His writings (at least through the 1970s) were of two types: sociological and theological (as a committed Protestant). On the one hand, he pursued social science studies of how various aspects of the world appeared from a strictly secular perspective; he then complemented these with studies examining the same phenomena from the point of view of the Bible. Biblical studies dealing with the challenges of a post-Christian secular world appealed to evangelical Americans struggling with 1950s materialism and 1960s cultural liberations. Neither approach gained much traction in the academic philosophical community, certainly not in the English-speaking world. In a paradox of contemporary intellectual culture, Ellul's sociological theory of technology achieved a cardboard cutout currency that allowed it to be easily dismissed.

When the secular study of *La Technique ou l'Enjeu du siècle* originally appeared in 1954, it received almost no attention outside the French evangelical community. A decade later in southern California, Aldous Huxley recommended to the Center for the Study of Democratic Institutions, a small think tank looking for projects, that the book deserved translation. In 1964 an English version appeared as *The Technological Society,* and in the lead-up to publication the Center sponsored a symposium, the proceedings of which were collected in a theme issue of *Technology and Culture*, the journal of the Society for the History of Technology. Ellul's written contribution to this 1962 conference (he did not attend personally), together with his book, became a touchstone in early ethical-political debates about the technological transformation of the social order.

Ellul's symposium paper, "The Technological Order," briefly summarized the thesis of his book, that *technique* has become what Durkheim would have called a "social fact" and Marx a "social relation" or force that influences behavior independent of any individual user. In the 1940s Ellul "grew more and more convinced that technology is the element that would have caught [Marx's] attention" (Ellul 2004, 27). "So [he] began to study Technique, using a method as similar as possible to the one Marx used a century earlier to study capitalism" (Ellul 1982, 176). The basic "characterology" of Technique reveals it to be artificial (i.e., human made), semi-autonomous with respect to other social institutions, self-determining, expanding on its own terms, and constituted by a tightly interwoven linkage of means. "Technique has become the new *milieu*, all social phenomena are situated in it" (395).

Ellul's concept of *Technique* (sometimes written with a capital T) is challenging. He does not write about *technologie*, which in French refers to the study of techniques. Instead, he wants to talk about a special mode of practice that has become a new, dominating social phenomenon, as signaled by the capital T (the same way in French it is common to refer to the "State" as a phenomenon at a higher level of abstraction than a "state"), which is as present in management as in engineering. A comparison can be made with what another social scientist, George Ritzer, calls "McDonaldization": the strong tendency of "the principles of the fast-food restaurant [efficiency, calculability, predictability, and control] . . . to dominate more and more sectors of American society as well as of the rest of the world" (Ritzer 1993, 1). In his book, Ellul distinguished between technical operations (the plethora of technical skills that people have always used to do specific things) and the technical phenomenon (the integration of differentiated skills into a system); the former does not create a technological society, the latter does—as in the way that the principles of McDonaldization have defused into politics, sports, religious services, and the multiple network relationships of supply chain capitalism.

A six thousand-word appendix to "The Technological Order" adds to his original characterology an argument for the ambiguity of technical progress and exhibits well Ellul's typically mid-level social theoretical method. He eschews the label "philosopher," yet his social theory occupies a possibility space more grounded than the speculations of British idealism (which G. E. Moore and Bertrand Russell analytically buried in the early twentieth century) and more abstract than strictly data-based social science behaviorism (with its offshoot into experimental philosophy). Although he did no original empirical research, Ellul drew extensively on a wealth of research by others, as well as on common-sense experience available to any reflective participant in the mid-twentieth century European lifeworld, so as to abduct concepts that could facilitate critical cultural self-examination.

In this spirit, the core of his essay sought to place the emergent phenomenon of *technique* into larger philosophical perspective by examining multiple observations about technological society current at the time. In the process, Ellul first identified what he saw as "fake problems": disagreeable features such as urbanization and pollution, the weakening of morals (in the sense of received social conventions), the sterilization of art, and the diminution of emotional life. Ellul thought Technique would eventually be able to solve these problems. The solution to such (fake) problems of technology is simply more technology. Technology can clean up the pollution it causes. It will become a platform for new social conventions and forms of art. Emotional life will find new expressions.

A second, central section, spells out what Ellul considered the real problems posed by technological development in terms of a double question: Is the human being able to remain *sujet* in a world of proliferating means? "Can a new civilization appear inclusive of Technique?" (398).

Interpreting the first version is complicated by lack of a French text. The translator gives, "Is man able to remain master in a world of means?" but notes that "master" is his rendering of the French *sujet*, since "subject" would be wholly misleading. Ellul is not asking whether humans can remain subject to technique. Ellul's elaboration, however, renders "master" itself inadequate. For Ellul, the uniqueness of human history has been achievement of a subjectivity that experiences itself as not wholly subordinate to its milieu, whether natural or social. The question is whether this subjectivity, this liberty, can be retained and cultivated in a new technological milieu. The question is whether the human is "capable of giving direction and orientation to Technique" (399). According to Ellul, although philosophers, engineers, scientists, and politicians unite in proclaiming the importance of values, these values either are presented to justify what already is or "are generalities without consequence" (399).

The 1980s sociological program to disclose the social construction of science and technology radically rejected Ellul's claim as no more than its own generality without empirical basis. People are socially constructing and deconstructing technologies all the time. But is it not possible, half a century later, to see in the trajectory of technocultural development some truth in Ellul's questioning? Precisely where is the human mastery being exercised in what Bruno Latour (2017) has described as our environmental mutation?

The second question turns on the contested idea of civilization. For Ellul, as for Arnold Toynbee Jr. (1934) and Norbert Elias (2000), civilization is a complex social organization that regulates ("restrains," for Elias) human behavior in some distinctive way. We can speak of synchronic contrasts, for example, of European (Christian) and Chinese (Confucian) civilizations. Samuel Huntington (1996) theorized a global "clash of civilizations," especially between Western, Islamic, and Confucian civilizing frameworks. For Ellul, however, a more illuminating contrast is diachronic between milieux of nature (hunters and gatherers), society (based in the domestication of plants and animals), and technique (since the Industrial Revolution).

This new technical milieu is problematic insofar as it creates a material culture that knows little beyond aggregate growth in power and productivity eventuating in global consumerism with an individualist market overlay. Technique tends to undermine the qualified autonomy from all that might seek to determine me (whether nature or society) while accepting and acting in awareness of my frail, finite but open facticity. Ellul nevertheless rejects any simple return to the past.

> Our duty is to occupy ourselves with the dangers, errors, difficulties, and temptations [of the present]. . . There is no possibility of turning back, of annulling, or even of arresting technical progress. What is done is done. It is our duty to find our place in our present situation and in no other. Nostalgia has no survival value in the modern world and can only be considered a flight into dreamland. (403)

Analytically one can identify two paths for the exercise of this duty. One believes "that the problem will solve itself"; the second thinks that "a great effort or even a great modification of the whole human being" is demanded (403). The former is the ideology of politicians, scientists, engineers, and economists who commit themselves to accelerating the process. The latter is found among philosophers, of whom he mentions Albert Einstein, Henri Bergson, and Pierre Teilhard de Chardin. Ellul himself is sympathetic to Bergson's (1932) argument for and yet skeptical about a possible *supplément d'âme* (supplement of soul) to rebalance the proliferation of technical powers. Unable to conceive of any concrete program, Ellul simply itemizes five "necessary conditions for a possible solution" (408): a correct diagnosis, an attack on the mythology of *technique*, cultivation of detachment from technology, critical reflection, and sustained interaction with scientists and engineers.

For Ellul the demythologizing of *technique* is ultimately a spiritual task. Parallel to his sociological studies, beginning with *Présence au monde moderne: Problèmes de la civilisation post-chrétienne* (1948), Ellul undertook a series of complementary theological analyses of technology from a biblical perspective. Over the course of five working decades, Ellul pursued this dialectical (sociological and theological) approach in more than 50 books to produce what is arguably the most extended critical engagement with modern technological civilization.

Unlike Turing's piecemeal analysis of specific issues, Ellul represents a European tradition of broad social philosophical criticism in which, however, he was the first to make technology the central theme. His tradition can be traced back to Rousseau's, Marx's, and Durkheim's efforts to analyze the pathologies of modern social institutions, especially the problematics of a social order in which religious affiliations have become attenuated or deeply distorted. Ellul presents technology as a new dominating presence that invites contestation from a biblically informed social theory resting on the Christian radicalism of Søren Kierkegaard and Karl Barth. In this explicit appeal to the teaching of the Bible as a counterpoint to technological power, of revelation over against reason, Ellul could not help but marginalize himself even more than by his choice of a professional life in Bordeaux over one in Paris.

Within philosophy of technology, Ellul's aspiration never sprang to professional life after the manner of Turing's analyticism. Even though thinkers as disparate as Lewis Mumford, Herbert Marcuse, Ivan Illich, Donald Verene, and Langdon Winner could be described as exhibiting affinities, their work is largely dismissed in a world intellectually flattened by commitments to innovation and technological change. The extent to which the post-social constructivist Latour and the provocational Peter Sloterdijk ignore Ellul remains somewhat inexplicable.

4. The Question Concerning Heidegger and Technology

Heidegger is undoubtedly the most consequential of the three contributors to classic European philosophy of technology considered here—and the most controversial. His *Being and Time* (1927) created a revolution in phenomenology and is recognized as one of the great works in German if not European and world philosophy. The controversy is that Heidegger joined and actively supported the Nazi Party, and in posthumous publications espoused anti-Semitism. This has raised a question about whether Heidegger's philosophy as a whole, with its argument for time as the horizon for understanding the meaning of being, deprives humans of a basis for judging historical actions and is thus fundamentally nihilistic.

For Heidegger, however, nihilism is manifest in modern technology insofar as technology is accepted as a kind of fate. "The Question Concerning Technology" (the 1977 translated title of "Die Frage nach der Technik," published in 1954) makes the argument in three roughly equal steps. It begins by questioning the commonsense definition of *Technik* or technology as a means created by humans to achieve some particular end. Although this instrumental or anthropological account may be descriptively correct, "the essence of technology [that is, what is most fundamentally taking place in its creation of instrumentalities] is by no means anything technological" (Heidegger 1977, 4). As an approach to this essence Heidegger takes a detour through Aristotle's account of how four causes conspire to bring-forth entities and distinguishes two basic modes of becoming: *physis* (nature) and *poiesis* (poetry or art). The sprouting and growth of an acorn brings an oak tree into the world. The *techne* or craft of an artisan takes wood from the oak and makes (*poiein*) a bed.

There is nevertheless more going on than just the particular outcomes of these two modes of becoming. Each mode differentially reveals or discloses reality, that is, is a type of *aletheia* (*a-letheia* or "un-hiding"), a Greek word rendered in English as "truth." "Technology is no mere means [but] a way of revealing" (12). *Physis* reveals the dynamic vitality of independently emerging entities that engenders an interwoven order; the truth of nature is *cosmos*. *Poiesis* reveals the hospitableness of nature to a human presence in the cosmos; the truth of technics is dwelling or inhabiting.

Against this background Heidegger takes a second step and asks further: What type of truth is happening in the historically distinct form of making that is modern technology? *Technik* is not the same as *techne*. Whereas *techne* works in the first instance, as in agriculture, to cultivate nature or assist its independent bringing forth of things that sustain human dwelling, and in a second instance, as in craft, to give to naturally occurring materials and energies supplementary forms especially commodious to human flourishing, modern technology imposes itself on nature. “The revealing that holds sway throughout modern technology does not unfold into a bringing-forth in the sense of *poiesis* [but] is a challenging [*Herausfordern*]” (14).

> The revealing that rules through modern technology has the character of a setting-upon, in the sense of a challenging-forth. That challenging happens in that the energy concealed in nature is unlocked, what is unlocked is transformed, what is transformed is stored up, what is stored up is, in turn, distributed, and what is distributed is switched about ever anew. Unlocking, transforming, storing, distributing, and switching about are ways of revealing. (16)

In a word, what modern technology discloses is reality as *Bestand*, resource, something able to be manipulated at will. This is the sense in which Heidegger argues technology is nihilism. It is a power cut loose from the restraints and guidance traditionally inherent in *poiesis* and *techne*, which must always acknowledge some measure of subservience to nature and human inhabitation. Premodern technics was (and remains) embedded in nature. Modern technology turns the tables and embeds nature in itself; modern technology constructs its nature.

Heidegger’s characterization of technoculture as a destruction of the craft lifeworld ontologizes a criticism of modernity variously expressed by, among others, theologian Romano Guardini (1927) and poet Friedrich-Georg Jünger (1946). For Heidegger what is taking place is not simply some human activity. Modern challenging-forth technology that reveals the world as resource itself arises from a “challenging that sets upon humans to order the real as resource” (19). His name for this challenging that animates human engagement in modern history as technology is *Gestell* (commonly translated as “Enframing”). *Gestell* is the true essence of modern technology which is not itself anything technological.

In a third, final step, Heidegger considers whether *Gestell* constitutes modern technology as fate. This is the big question concerning Heidegger and technology. “The essence of modern technology [initiates a] revealing through which the real everywhere, more or less distinctly, becomes *Bestand*” (24). But this human destiny “is never a fate that compels” insofar as we become aware of what is taking place and thereby take up a free relationship to it (25). “In this way we [sojourn] within the open space of destining, a destining that in no way confines us to a stultified compulsion to push on blindly with technology or, what comes to the same thing, to rebel helplessly against it and curse it as the work of the devil” (25–26). An uncanny contradiction in Heidegger was his choice of the violence of National Socialism as a way not to push blindly on or to helplessly rebel.

All revealing brings with it danger, and is in all of its modes necessarily danger. The revealings of *physis* and *poiesis* both endanger more primal senses of what is; they tempt us to think of reality as no more than a system of causal relationships or of beauty as founded in human art. *Gestell* as a new way of revealing presents a still greater danger: It tempts human beings to think of themselves simultaneously as *Bestand* and as all powerful. It becomes increasingly difficult for us to think in any other way. The threat from technology "does not come in the first instance from the potentially lethal machines and apparatus." Instead, G*estell* threatens humans with the possibility of being unable "to enter into a more original revealing and hence to experience the call of a more primal truth" (28). As this danger grows, however, as with all dangers, Heidegger affirms, quoting from the poet Friedrich Hölderlin, there also grows a "saving power."

This oracular quotation can have multiple interpretations. In a psychological sense, it is true that when caught in a bind we are often able to imagine courses of action that eluded us in more relaxed circumstances. A more deeply provocative suggestion in the text is that this very essay and Heidegger himself may be the saving power emerging in our fraught time. Whereas Ellul drew inspiration from Kierkegaard's radical Christian criticism of secular culture, Heidegger echoes Friedrich Nietzsche's aggrandizing conception of philosophy. Maybe "only a god can save us," but the god needs a prophet. The posture is one that has spilled over into many of those influenced by Heidegger, even as they profess to think against him.

On less grandiose levels, Heidegger's problematizing of technology in phenomenological terms has been and remains animating of multiple practices. His influence, like Turing's, remains seminal. One cannot fully appreciate the work of major post-classical philosophers of technology as different as Hans Jonas, Albert Borgmann, and Don Ihde without some attention to the living influence of Heidegger. However, as with Turing, explicit ethical and political philosophical dimensions are dormant in Heidegger and have only been midwifed to phenomenological birth by others.

5. In Defense of the Dead

The three texts used here to exemplify classic European philosophy of technology have superficial similarities. They range in length from nine to 12 thousand words and are essay-like. A google n-gram shows that from 1950 well into to the 1980s the names and ideas of Turing, Ellul, and Heidegger were all increasingly referenced in English language books. It is not possible to tell from the n-gram the extent to which the referencing was associated with an emerging philosophy of technology, but it is reasonable to suspect that, with regard to Turing (if we include philosophy of computing as philosophy of technology) this was very often the case; with Ellul, it might have been half-and-half (given his bimodal influences); and with Heidegger, it was probably a minor element (given Heidegger's defining twentieth century presence in so many philosophical projects). Since the 1990s, references to all three first tapered off and then slightly declined; with Ellul the decline was most precipitous, no doubt reflecting a culturally dominant idolization of technology and the sustained rejection of his allegedly monolithic view and pessimistic determinism. Referential persistence to such contributors to philosophical reflection on technology, if nothing else, suggests that any full appreciation of historical developments in the field and its current scope would benefit from at least minimal acquaintance with all three.

In this respect, we can notice how Turing almost single handedly initiated a professional tradition of analytic work on computing and information technology that has expanded into active research in the philosophy of mind and cognitive psychology. His model of engaging not with some monolithic technology but some particular technology has been further projected into an expanding spectrum of living philosophical engagements, from nuclear weapons to technomedical advances and nano-engineering. Turing's pro-attitude toward if not celebration of computers has broadly prevailed over any skeptical stance, even when analysis has been enlivened by attention to ethical dimensions slighted in the original. Stephen Toulmin's (1982) observation about how this played out in philosophical reflection on biomedical technologies makes for useful analogy. Computer ethics has an often slighted legacy in Wiener's (1950) less sanguine reflections on the social implications of cybernetics, which deserves to be revisited in an era of big data AI and algorithmic capitalism.

Ellul presents a far more problematic sociological theory of technology implicating ethics and politics at the level of socio-historical roots. In his large body of work it is difficult not to read a paradoxical retreat in action from global to local folded in with a leap of faith into the absolute. "Think globally, act locally," was his secular motto; the primacy of Christian revelation, a sacred belief. Absent the rhetorical cleverness of many French theorists, such an absurd combination has not unexpectedly been relegated to the margins of viability. Modest exceptions highlight only shallow breaths of his thought, as in Langdon Winner's (1977 and since) persistent gadfly questioning of pro-technology ideologies and the occasional reference to Ellul's analysis of the techniques of propaganda (1965) in media and communication philosophy.

Heidegger undertook a phenomenological reflection on technology leading to ontological claims which, despite a manifold of rhetorical attractions has continued to be philosophically fruitful; even when its ontological dimensions have withered, phenomenologies of technology continue to sprout. Heidegger's Nazism nevertheless remains a fundamental stumbling block for which almost any reference must apologize. In regard to technology, Heidegger now functions as a philosopher to think against as much as with. His presence remains in attacks against him.

These three early models of analytic, social-political, and metaphysical philosophy of technology in their differential presence return us two simple questions: To what degree should either or all simply be referenced and left behind? Is there any sense in which together in their combined legacies they might suggest something hidden amid the works tumbling out of our bookshelves and now swamping the document folders of our cloud storage?

More important than either similarities or contrasts is the ease with which it is possible to engage the relevance of specific arguments in Turing and to recognize the continuing vitality, indeed dominance, of analytic practices in the philosophy of technology versus the difficulties in discovering robust living continuations of the core features in Ellul and Heidegger. With Ellul and Heidegger, not only is it difficult to isolate particular arguments that can be subjected to conceptual refinement or critical assessment; their claims seem so large and consequential that to practice philosophy in their wake in any academic setting appears foolish if not impossible. Small scale, piecemeal thinking about well-focused particulars is the only academically viable option when addressing technology. Take the example of ethics: The ethics of industrialization, of nuclear power, of biomedical and biological technologies, of computers and information technology, of genetic engineering, of nanotechnology, of climate change, and of robotics are each of them so hard that it seems irrational to even imagining thinking the ethics of technology as a whole. The best that seems feasible is establishment of interdisciplinary centers for something more circumscribed such as biomedical or engineering ethics. In their ambitions, Ellul and Heidegger set themselves up for being ignored if not rejected. Philosophical life today virtually *requires* that they be (perhaps monumentally) entombed. And yet ...

And yet taking a broad and inclusive look back over the trajectory of philosophical encounters with technology, was it not the big thinking of such figures as Rousseau, Bentham, Owen, and Marx that helped stimulate and guide beneficial social change? Could industrial capitalism have been even as partially reformed as it was during the late 19th and 20th centuries without some stimulus from classic big philosophical social theories and ideas that would be classified as conceptually fuzzy by more rigorously thinking analytic philosophers? Would the Limited Nuclear Test Ban Treaty have been possible, absent some dramatic large-scale criticisms of nuclear weapons? Would agencies of technology assessment have been established without multiple general criticisms of the unintended consequences of technoscience and a nascent anti-technology movement? Would the US Environmental Protection Agency have been created in the absence of the big (if less than professionally philosophical) thinking by Aldo Leopold and Rachel Carson? Did Ellul's criticisms of the technological society not provide intellectual encouragement for opposition to America's technological war crimes in Vietnam?

The piecemeal approach to thinking about technology exhibits distinctive parallels to neoliberalism. Just as neoliberalism declares, in Margaret Thatcher's famous words, "There's no such thing as society," empirical turn philosophers of technology seem to imply there is no such thing as Technology with a capital T. The social ontology of neoliberalism finds a natural ally in what might be called a neoliberal philosophy of technology, a philosophy that can leave the techno-lifeworld to be socially constructed by captains of engineering and innovation under little more than *ex post* attention to safety here, privacy there, and distributive justice adjudicated by marketplace rationality.

Is it not the case that whenever big thinking is rejected in favor of addressing manageable problems, it implicitly rests on a comfortable affirmation of the status quo? Latent in the manifest commitment to analytic meliorism, is there not an unspoken commitment to things as they are, including a measure of existential pleasure in the engineering away of respect for human and planetary life as it has been known for thousands of years? Under such circumstances, is it not necessary at least on occasion to think big again?

6. Coda

Then there are the results of the modest personal pedagogical experiment in which this essay incubated. Devoting a graduate seminar in China to three imputed classics in European philosophy of technology generated a unique level of engagement—and, paradoxically, a small opening, however limited and provisional, to classical Chinese philosophy. Heidegger and Ellul both argue that modern technology owes something to the Western philosophical heritage. Acknowledgement of problematic consequences in technology cannot help but invite consideration of alternatives. Just as the first wave of Chinese modernization in the form of enthusiasm for modern science and technology during the late 1800s and early 1900s sponsored radical criticisms of Chinese traditions (Chen Duxiu's [1919] proposal for replacing "Mr. Confucius" with "Mr. Science and Mr. Democracy"), so today a second wave of modernization may stimulate reflective reconsiderations of Daoism, Confucianism, and Buddhism (as in Gan Yang's [2007] call for "Confucian socialism").

For myself as well, the seminar re-convinced me that Ellul's questions merit being exhumed from an analytic or social constructivist tomb. In a world simultaneously engulfed by increases in techno-fragilities, global environmental threats, and fantasies of anti-global nationalism, Ellul deserves to be treated as more than a zombie and may well serve as a tincture antidote to the fast-paced celebrity culture of academia in which intellectuals compete with one another to coin captivating terms and philosophy struggles to keep up with itself. Especially in relation to technology, philosophers too often seem at pains to one-up each other with flashy arguments that on careful examination add little genuine insight or guidance. A return to classic texts can serve as salutary counterfoil.

Although Ellul argues against nostalgia and any attempt simply to return to the dreamland of the past, there is an even more seductive dreamland of techno-fiction fantasies about the infinite benefits and even necessity of innovation forever. Under such conditions, surely there is some good in trying to think not only with technology but against it.

References

Achterhuis, Hans, ed. 2001. *American Philosophy of Technology: The Empirical Turn*. Trans. Robert P. Crease. Bloomington, IN: Indiana University Press.

Bergson, Henri. 1932. *Les Deux Sources de la morale et de la religion*. Paris: Félix Alcan. English trans. R. Ashely Audra and Cloudesley Brereton, *The Two Sources of Morality and Religion* (London: Macmillan, 1935).

Chen, Duxiu 陈独秀. 1919. Xin qingnian zui'an zhi dabian shu 新青年罪案之答辩书 (Reply to criticisms of *New Youth*), 6, no. 1 (January 15): 1–2.

Elias, Norbert. 2000. *The Civilizing Process: Sociogenetic and Psychogenetic Investigations*. Rev. from German, 1939. Trans. Edmund Jephcott. Malden, MA: Basil Blackwell.

Ellul, Jacques. 1948. *Présence au monde moderne: Problèmes de la civilisation post-chrétienne*. Geneva: Roulet. English trans. Olive Wyon, The Presence of the Kingdom (Philadelphia: Westminster, 1951). New English trans. Lisa Richmond, *Presence in the Modern World* (Eugene, OR: Wipf and Stock, 2016).

Ellul, Jacques. 1954. *La Technique ou l'Enjeu du siècle*. Paris: A. Colin. English version, John Wilkinson, *The Technological Society* (New York: Random House,1964).

Ellul, Jacques. 1962. "The Technological Order." *Technology and Culture*, 3, no. 4 (Autumn 1962): 394–421.

Ellul, Jacques. 1982. *In Season and Out of Season: An Introduction to the Thought of Jacques Ellul*. Trans. Lani K. Niles. San Francisco: Harper and Row.

Ellul, Jacques. 2004. *Perspectives on Our Age: Jacques Ellul Speaks on His Life and Work*, edited by Willem H. Vanderburg. Rev. ed. Toronto: Anansi.

Gan, Yang 甘阳. 2007. *Tong san tong* 通三统 (Synthesizing three traditions). Beijing: SDX Press.

Guardini, Romano. 1927. *Briefe vom Comer See*. Mainz: Joseph Weiger. English trans. Geoffrey W. Bromiley. *Letters from Lake Como: Explorations on Technology and the Human Race*. Grand Rapids, MI: Eerdmans, 1994.

Heidegger, Martin. 1954. "Die Frage nach der Technik." In *Vorträge und Aufsätze*. Pfullingen, Germany: Neske, 13–44. English trans. by William Lovett, "The Question Concerning Technology." In *The Question Concerning Philosophy and Other Essays*. New York: Harper and Row, 1977, 3–35.

Hofstadter, Douglas R. 1979. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Random House.

Huntington, Samuel P. 1996. *The Clash of Civilizations and the Remaking of World Order*. New York: Simon and Schuster.

Ihde, Don. 1993. *Postphenomenology: Essays in the Postmodern Context*. Evanston, IL: Northwestern University Press.

Ihde, Don. 2018. "Should Philosophies Have Shelf Lives?" *Journal of Dialectics of Nature*, 40, no. 1 (January): 98–104.

Jünger, Friedrich-Georg. 1946. *Die Perfektion der Technik*. Frankfurt am Main, 1946. English trans. Frederick D. Wilhelmsen. *The Failure of Technology: Perfection without Purpose*. Chicago: Regnery, 1949.

Kroes, Peter, and Anthonie Meijers, eds. 2000. *The Empirical Turn in the Philosophy of Technology*. Research in Philosophy and Technology, Vol. 20. New York: JAI Press.

Latour, Bruno. 2017. *Facing Gaia: Eight Lectures on the New Climatic Regime*. Translated by Catherine Porter. London: Polity Press.

Lucas, J. R. 1961. "Minds, Machines and Gödel." *Philosophy*, 36: 112–127.

Mitcham, Carl. 1993. "How *The Technological Society* Became More Important in the United States than in France." In *Jacques Ellul and the Technological Society in the 21st Century*, edited by Helena M. Jerónimo, José Luís Garcia, and Carl Mitcham, 17–34. Dordrecht: Springer.

Penrose, Roger. 1989. *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*. Oxford, UK: Oxford University Press.

Ritzer, George. 1993. *The McDonaldization of Society: An Investigation into the Changing Character of Contemporary Social Life*. Newbury Park, CA: Pine Forge Press.

Toulmin, Stephen. 1982. “How Medicine Saved the Life of Ethics.” *Perspectives in Biology and Medicine*, 25, no. 4 (Summer): 736–750.

Toynbee, Arnold J. 1934. *A Study of History*. Oxford: Oxford University Press.

Turing, Alan. 1950. “Computing Machinery and Intelligence/” *Mind*, New Series, 59, no. 236 (October): 433–460.

Wiener, Norbert. 1950. *The Human Use of Human Beings*. Cambridge, MA: MIT Press.

Winner, Langdon. 1977. *Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought*. Cambridge, MA: MIT Press.

CHAPTER 3

THE EMPIRICAL TURN

PETER-PAUL VERBEEK

1. Introduction

The history of the philosophy of technology is marked by a transition in its approach to technology. This transition is often indicated as an 'empirical turn' (Achterhuis 2001; Kroes and Meijers (eds.) 2001). After a long period of broad philosophical reflection on 'Technology' and its impact on 'Society,' scholars in the field started to feel discomfort with this generalizing and sometimes rather monolithic approach to technology. A more differentiated perspective was needed, that was able to do justice to the differences between various types of technologies and their social implications. The need to change perspective was fed by the advent of the empirical field of Science and Technology Studies (STS), in which disciplines like sociology, anthropology, and history joined forces to investigate science, technology, and their interactions with society. In the philosophy of science, STS had already resulted in an empirical turn: scholars started to analyze science as a *practice* and not only as *theory*.

On this basis, in the 1980s the philosophy of technology also started to develop closer connections to the concrete, empirical reality of technological artifacts, systems, and processes. In order to analyze technology in a better way, many philosophers decided that their starting point should not only be the philosophical tradition, but also a closer understanding of technology itself, and its actual relations to human beings and society. This turn, obviously, did not imply that the philosophy of technology became an empirical discipline itself. Rather, philosophical reflection started to look for new ways to engage with the phenomena it aims to understand philosophically and evaluate ethically.

In this chapter, I will discuss this empirical turn in four steps. First, I will sketch the main lines of thinking in 'classical philosophy of technology' that gave rise to the need for a reorientation. Second, I will discuss what the empirical turn entailed, and which forms it took. After this, I will focus on the impact the empirical turn had on the ethics of technology, including the new ways it opened for ethics to engage with technological practices themselves. And finally, I will discuss how the field is developing *after* the empirical turn, and how the concept of a 'turn' keeps inspiring scholars to innovate the field.

2. Classical Philosophy of Technology

The empirical turn in philosophy of technology emerged from one of the central debates in the field: the question to what extent we should see technology as a neutral tool or as a determining force in society. Is technology ultimately an instrument in the hands of human beings, deriving all its impact from human actions and intentions, or does it actually have the capacity to change society beyond, and perhaps even independently from, human intentions? These two positions are often referred to as 'instrumentalism' and 'substantivism,' respectively (Borgmann 1984).

The instrumentalist approach views technology as just a means to an end, a neutral tool to achieve human goals. How could technology be more than an instrument? Since technology is not able to set itself any goals, it is always dependent on human beings for its development, implementation, and use. Seeing technologies as more-than-instrumental would downplay or ignore the responsibility that human beings ultimately have for technologies (Pitt 2014). The substantivist approach takes exactly the opposite position. It considers the claim that technology is neutral to be misleading: instrumentalism downplays how technology has in fact always changed society, and often in different ways than its designers intended. In order to understand and evaluate technology and its relations to society, therefore, we need to take technology seriously as a substantive force in culture and society.

In this debate about technology and society, philosophical and empirical claims go hand in hand: philosophical conceptualizations of the *relations* between technology and society are always connected to assumptions about their actual *interactions*. On the one hand, philosophy of technology aims to develop theoretical concepts to characterize technology and its relations to society. But on the other hand, it inevitably bases itself on upon empirical claims about technology and its actual interactions with society. This ambiguity made an empirical turn in the field almost inevitable. As I will argue, the dominant role of the substantivist approach resulted in a growing discomfort with the implicit or explicit empirical claims it made, which paved the way for a reorientation on how to engage with technology and its relations with society.

2.1 Substantivism and Its Critique

Substantivism has milder and stronger variants. While many contemporary philosophers of technology do recognize the non-neutrality of technology by analyzing its impact on human beings and society, only very few of them would subscribe to a strong substantivist position. This position considers technology to be both deterministic and autonomous*:* it believes that technology plays a determining role in culture and society, and that technology develops in ways that cannot be controlled by human beings.

In the early positions in philosophy of technology, the strong substantivist view was quite dominant. Many authors were concerned that 'Technology' was running out of control and started to change society in irreversible and negative ways. And some of them urged that we should regain sovereignty over technology, breaking its determining role and taking it back to its original instrumentality. The French thinker Jacques Ellul, for instance, has argued that technology should be seen as a "system" that transforms our entire society: it introduces a framework of efficiency and effectiveness that becomes the new background against which we interpret the world, and in which non-technological values and phenomena play an ever less significant role (Ellul 1964). Moreover, he claimed, the technological system is self-affirming: it results in a "technical fix," which seeks to solve all problems generated by technology with new, more effective and efficient technologies. There is no escape from technology: every obstacle the system meets results in an expansion of the system rather than a move away from it.

The strong substantivism of the early positions in the philosophy of technology, which are now often indicated as 'classical philosophy of technology,' typically went hand in hand with a quite pessimistic approach. Many of these early positions developed from Marxism or from phenomenological theories, and focused on the forms of *alienation* that technology can bring about: alienation from nature, but also from ourselves as human beings.

In order to get a closer understanding of this specific combination of determinism, autonomy, and pessimism, the work of Martin Heidegger can serve as an example, being one of the most prominent 'classical' thinkers in the history of philosophy of technology. To understand technology, Heidegger claimed, we should not see it as something instrumental, or something made by human beings. Technology is much more than an instrument: it embodies a way of *understanding the world*. Moreover, human beings cannot choose for this way of understanding the world: it belongs to a historical development that we are all part of. To understand what Heidegger meant by this, we need to place Heidegger's approach to technology in the context of his philosophy of 'being.' In Heidegger's view, human thinking is always guided by a fundamental understanding of what it means 'to be,' and this understanding develops over time, beyond human control. Where for Medieval philosophy 'being' meant "having been created by God" and for modern philosophy "being an object in relation to a subject," modern Technology turns 'being' into "being raw material for the human will to power." The reality of things has come to consist in what humans can make of them.

As expression of this 'will to power,' modern Technology changes reality in a stockpile of resources. While an old, wooden bridge over the river Rhine, in one of Heidegger's well-known examples, still recognized the river Rhine in its own right, a water power station forces the river to show itself as a supplier of energy (Heidegger 1977, 16–17). According to Heidegger, this fundamental understanding of the world as raw material, available for human manipulation, results in a dangerous situation. In fact, he considered Technology to be "the greatest danger" (Heidegger 1977, 27). This danger lies in the fact that technology offers no escape from its highly limited and narrowed-down understanding of the world. Every attempt to work toward an alternative way to understand the world would itself be a technological act: an attempt to exert power over the fact that we are exerting power. Climbing out of the technological framework to understand the world immediately throws us back into that framework. The only thing we can do, Heidegger claimed, is to develop an attitude of "releasement." From this attitude, we could develop a paradoxical "will not to will," recognizing the technological character of our time, without being determined by it, in order to stay open for the development of a new way of understanding 'being.'

While Heidegger's work is still influential, it has been also been sharply criticized along lines that illustrate the central dimensions of the empirical turn. Political philosopher Andrew Feenberg, for instance, has criticized Heidegger's approach for its monolithic character: since it does not make any distinction between different types of technology, it does not have much to offer to scholars who want to engage with actual technologies and their social implications (Feenberg 2000). Don Ihde especially criticized Heidegger's romantic preference for old technics over modern technology. According to him, Heidegger fails to see that older technologies can also exert power over nature, while modern technologies also have the potential to bring us closer to nature. In these critiques, a dissatisfaction becomes visible with Heidegger's lack of connections to actual technologies and their implications for human beings and society (Ihde 1993). The category of 'Technology' (with a capital T) appeared to be too broad to grasp the subtlety of human-technology relations; the social impact of technology might be less deterministic than the 'history of Being'; and Heidegger's overly pessimistic image of technology needs to be replaced with a more nuanced and ambivalent one.

2.2 Transcendentalism and Beyond

This critique of Heidegger's philosophy of technology illustrates the growing resistance of a new generation of thinkers against the classical positions. The monolithic and pessimistic character of classical philosophy of technology appeared to belong to the specific historical context in which it developed. The rapid and radical processes of social change that resulted from industrialization and automation resulted in feelings of alienation. But over the course of time, the radical opposition between humans and technology became less convincing for many philosophers. First of all, the pessimistic character of the classical positions did not do justice to the positive contributions that technology made to society as well. And second, the deterministic character of the classical analyses started to be at odds with the growing body of research of the social dynamics of technological developments in the empirical field of Science and Technology Studies, which showed that technology does not *determine* society, but is itself socially *constructed and appropriated.*

Gradually, therefore, technology started to be understood as an *element of* culture and society, rather than being *opposed to* it. While the classical positions in the field approached technology as alien to human beings and as a potential threat to the authenticity of human existence and human's understanding of the world, new positions started to question this opposition. Instead of saving culture from technology, it also appeared to be possible to study technological culture,' as will be explained in the next section.

As I have argued elsewhere (Verbeek 2005), classical analyses of technology typically followed a 'transcendentalistic' approach: they analyzed technology in terms of its *conditions*. Heidegger's work follows this pattern as well: his approach to technology as a way of understanding the world in fact reduced technical devices and systems to the way of thinking behind them. He did not analyze technology itself, but the technological *thinking* from which it originates. In his view, in the end, technologies do not *produce* our will to power, but are rather the *result* of it. Even in his example of the water power plant in the river Rhine, it is ultimately not the power plant that makes humans understand the river as a source of energy—rather it is the technological approach to nature as raw material that made it possible for us to develop water power plants in the first place (Verbeek 2005). The 'essence' of technology—to phrase it in a Heideggerian way—is not in the technological artifacts themselves but in the overpowering way of understanding the world behind them.

It is exactly this transcendentalism that is abandoned in the empirical turn. Rather than reducing technological artifacts, systems and practices to the conditions that lie behind them, it started to take them as a *starting point*. Empirical insights in human-technology relations, design and innovation processes, and the social implications of technologies became a central element of philosophical analysis.

3. The Empirical Turn

The empirical turn, which started in the 1980s, reversed the perspective of classical philosophy of technology. While classical positions in the field tended to reduce technological artifacts, systems and practices to their *conditions*—like the technological way of thinking behind them, or the system of mass production that they are part of—the empirical turn urged philosophers of technology to take the empirical reality of technology as a *point of departure*. And rather than making claims about "Technology" as a broad social and cultural phenomenon, philosophy of technology started to focus on actual *technologies*, in their concrete contexts. This turn toward concrete technologies and practices took two directions, one focusing on the *social implications* of technologies, the other on *engineering practice*.

3.1 Technology and Society

The first variant arose particularly from North-American approaches to technology (see Achterhuis 2001; Brey 2010). Andrew Feenberg, for instance, integrates empirical work from Science and Technology Studies in his neo-Marxist and phenomenologically inspired approach of 'critical constructivism' (Feenberg 1991). This approach takes the mutual shaping of technology and society as a starting point, and in doing so it opens new perspectives on the relations between technology, power, and democracy. Rather than asking how 'the technological system' is intruding upon 'the human lifeworld,' as classical philosophy of technology would frame it, Feenberg asks how concrete technologies rearrange power relations, and how technologies and design processes can themselves be democratized.

For his analysis, Feenberg combines philosophical theory with insights from Science and Technology Studies, most notably from the constructivist approach of actor-network theory, which views technologies as both giving shape and being shaped by their social context (Latour 1987). Technologies are constructed in networks of relations, in which human actors play a central role, with their interpretations, interests, and ideas, but in which also technologies themselves play an active role as 'actants,' in the sense that they help to shape the networks of relations in which other entities are constructed. Feenberg's 'critical constructivism' aims to make visible the politics of these constructions by engaging actively and critically with the processes of construction themselves. Understanding the dynamics of technology development and of the interactions between technology and society opens up a new range of political questions regarding power resistance, inclusion, exclusion, and empowerment (Feenberg 1999).

A good empirical example of this political significance of technology can be found in the work of Langdon Winner, another empirically oriented North-American philosopher of technology. In his seminal article "Do Artifacts Have Politics?" (Winner 1986) he analyzed the example of lowly built bridges over the Parkways on Long Island, New York. The bridges, designed by architect Robert Moses, were allegedly built so low that only cars could pass below them, not buses. In the days these bridges were built, many African-American residents of New York City did not own a car, and "one consequence was to limit access to Jones Beach, a widely acclaimed public park" (Winner 1986). It is important to say that the veracity of Winner's interpretation is disputed (see also Joerges 1999, and Woolgar and Cooper 1999) but the example still shows that it is possible to approach technological artifacts as political entities, in this case as discriminatory, or even racist.

Another line of thinking in this new school of 'empirical philosophy of technology' is the so-called 'post-phenomenological' approach, initiated by North-American philosopher Don Ihde (see Ihde 1990; Selinger 2006; Rosenberger and Verbeek 2015). Postphenomenology explicitly moves beyond the romantic opposition of humans and technologies in classical phenomenology. Instead of criticizing Technology as a distortion of a more primordial or authentic human-world relation, it aims to understand how technologies *help to shape* the relations between humans and the world. And rather than locating technologies in the world of material objects as opposed to the world of human subjects, postphenomenology considers technologies to be part of the *relations* between human subject and the world of objects. Technologies do not only close off specific interpretations of the world, but also bring new ones: new social relations, new moral and aesthetic experiences, new scientific observations. From this perspective, technologies do not bring alienation but *mediation*.

When technologies are used, they bring about specific relations between users and their environment: users are typically not only interacting with the technology itself, but engage in a practice *via* that technology. When driving a car, for instance, people are not only interacting with the car itself, but also develop new types of behavior on the street and new experiences of the environment. And MRI scanners are not only complicated machines to interact with, but also help to shape how neuroscientists understand the brain and how neuropsychologists understand human behavior and perception (De Boer et al. 2020). Technologies mediate human-world relations, and help to shape the character of these relations, and people's understanding of the world (Verbeek 2015). They do so in many different domains, ranging from scientific practices to moral relations, and from existential questions to political engagement. In order to understand these mediations, we need to start from technologies themselves, rather than reducing them to their conditions.

3.2 Engineering Philosophy of Technology

Besides this societal variant, the empirical turn also has an engineering variant. Not the social implications of concrete technologies, but the concreteness of engineering practice has a central place then (Brey 2010). In all its attention to social implications, some scholars claimed, the philosophy of technology seemed to have forgotten to think about technology *itself*. Therefore, philosophers like Peter Kroes, Anthonie Meijers, and Joe Pitt have developed alternative accounts of technology, aiming to characterize technological artifacts, technological functions, and technological design (Franssen et al. 2016).

The 'dual nature' approach to technology that was developed by Peter Kroes and Anthonie Meijers is a good example of this variant of the empirical turn (Kroes and Meijers 2006). In their characterization of technology, they address the duality of every technical artifact, as being part of both the physical world of material objects and the social world of intentional subjects. On the one hand, they claim, technologies need to be seen as physical structures: material entities that follow the laws of nature; but on the other hand, these structures realize functions that are connected to human intentionality: functions are the outcomes of intentional design, and play a role in the realization of human intentions. This makes technological artifacts hybrid entities, requiring both a physical and an intentional conception of the world to be described adequately.

This duality repeats itself when we aim to understand technological artifacts as *functional* objects. Also here, two views can be distinguished: one related to the intentional approach of the world, the other to the physical approach. From an intentional perspective, technological functions are ascribed to objects by human agents, who embed the object in a means-end relation: in this relation, it becomes relevant as a means to achieve an end (Kroes and Meijers 2006). But from a physical perspective, technological functions are to be seen as the result of physical properties that contribute in a causal way to the capacities of the object (2006).

This notion of 'duality' or 'hybridity' plays a role in many other approaches in the philosophy of technology as well. In this respect, the engineering variant of the empirical turn has much in common with the social variant, even though the former is more closely associated with analytical philosophy and the latter with continental philosophy—if this distinction still holds in the 21st century. Both variants thematize the close connections between technology and society, and their conceptual and normative implications. Technologies connect the human world of 'subjects' and the physical world of 'objects,' and in doing so they challenge the sharp distinctions we often make between them.

In sum, the empirical turn has resulted in a more nuanced and open approach to technology than earlier approaches had, without giving up the critical impetus of classical philosophy of technology. On the one hand, the empirical turn has opened new directions to develop conceptual frameworks for analyzing technological artifacts and engineering practice. And on the other hand, the empirical turn has offered new conceptualizations of the interactions between human beings, technologies, and society, giving rise to new theories about the ethical implications of various technologies. In doing so, it has also made it possible to move from 'Technology critique' to 'ethics of technology,' as will be elaborated in the next section.

4. Ethics of Technology and the Empirical Turn

The empirical turn has not only made it possible to develop a closer *understanding* of technology and its relations to society, but has also had profound implications for its ethical *evaluation*. First of all, the closer understanding of the interactions between technology and society, as made possible by the empirical turn, gave rise to new forms of *applied ethics*, dedicated to the concrete ethical questions in specific technological fields, like engineering technology, information technology, and biomedical technology. This branch of the ethics of technology has resulted in various frameworks and approaches to address ethical issues related to technologies, e.g., regarding privacy, safety, security, and sustainability. But besides these new forms of applied ethics, the philosophy of technology also resulted in new *ethical theory*.

4.1 The Moral Significance of Technology

One of the central themes in the ethics of technology after the empirical turn has been the moral significance of technologies (Kroes and Verbeek 2014). Empirical-philosophical analyses of human-technology relations have shown that technologies play an important role in the moral actions and decisions of human beings, ranging from speed bumps that help to shape people's driving behavior (Latour 1999) to TV sets that shape how families spend their time in the evenings (Borgmann 1995) and sonograms that inform moral decisions about abortions (Verbeek 2008). This empirical observation raises the question to what extent technologies "have" ethics. Since ethics is about the question of how to act, and technologies help to shape human actions and decisions to act, there seem to be good reasons to see technologies as ethically charged. But how to conceptualize this moral significance of technology, in a philosophical discourse which connects ethics only to human subjects, not to technological objects?

Some authors actually attribute *moral agency* to technological artifacts. Bruno Latour, for instance, has proposed a 'symmetrical' approach to humans and nonhumans, in which both can be an agent, or 'actant,' as he prefers to call them (Latour 1993). Those who complain about a loss of morality in our society, he claimed, simply forget to look at things, and only look at humans, as the example of the speed bump illustrates. Humans do not have a monopoly on moral agency. Other authors are fiercely opposed to this attribution of moral agency to things. They argue that the essential conditions for agency, most notably the condition of intentionality, can never be met by things. Approaching technologies as moral agents would merely be a form of *anthropomorphism*: using human categories to speak about nonhuman entities to which these categories do not apply. Moreover, attributing agency to things could result in the idea that we would actually blame things for ethically problematic actions, which would reduce our sense of human responsibility (Peterson and Spahn 2010; Peterson 2012).

In order to avoid these two extreme views—and to do justice to both the philosophical hesitation to expand agency to nonhuman entities, and the empirical observation that technologies are nonetheless involved in moral actions and decisions—ethicists of technology have been developing an empirical-philosophical alternative. Rather than claiming that technologies "have" moral agency, they expanded the notion of moral agency in such a way that technologies can be *part of it* (Verbeek 2014) or *help to shape it* (Illies and Meijers 2009). Moral agency is not "in" technologies but comes about in the *interactions* between humans and technologies. From this approach, there is no need to attribute human characteristics to nonhuman entities (the 'philosophical' element of empirical philosophy), while still acknowledging that technologies do play a constitutive role in moral actions (the 'empirical' element in empirical philosophy.

A good example of this empirical-philosophical ethics of technology is the approach of 'moral mediation' (Verbeek 2011; Kudina 2019). From this approach, technologies play a mediating role in the moral relations that human beings are engaged in. Technologies-in-use mediate the relations between humans and the world (Verbeek 2005): they help shape how humans understand the world, and in doing so they also help to shape our moral decisions. Prenatal diagnostics, for instance, creates new moral relations between expecting parents and the fetus: because it makes it possible to anticipate the health condition of the child, it makes expecting parents responsible in new ways, and informs their decisions about parenthood and abortion (Verbeek 2008).

4.2 Value Dynamism and the Moral Appropriation of Technology

In recent work on the moral significance of technology, the notion of "moral mediation" was complemented with the notion of 'moral appropriation.' This line of research brings in another blend of empirical and philosophical investigation. In all our attention to the mediating role of *technologies* in moral relations, the actively interpreting role of *human beings* in human-technology relations remained underexposed. In the end, moral mediation is not only the result of the characteristics of mediating technologies, but also of the ways in which human beings interpret these technologies, and "appropriate" them as part of their relations with the world (Verbeek 2015). Ultrasound has the capacity to make the fetus visible, but this capacity becomes morally relevant only when it is appropriated as a possibility to anticipate the health condition of the fetus, and to take action on this.

Just like processes of mediation, processes of appropriation can be studied both empirically and philosophically. In her dissertation "The Technological Mediation of Morality: Value Dynamism and the Complex Interaction between Ethics and Technology" (Kudina 2019), Olya Kudina has developed a model for this empirical-philosophical study of mediated morality: the 'hermeneutic lemniscate.' Expanding the classical hermeneutic idea of the 'hermeneutic circle,' which explains how the interpreter and the interpreted constitute each other in acts of interpretation, her lemniscate model shows how this circular human-world relation is in fact mediated by technologies.

This technologically mediated hermeneutic circle connects human, technology, and world via a lemniscate, with the shape of the infinity symbol, ∞: humans interpret a technology (human—> technology), which then mediates human interpretations of the world (technology—> world); within this specific understanding of the world, the technology acquires a specific role and meaning (world—> technology), and constitutes the user in a specific way (technology—> human). In relation to ultrasound: humans use ultrasound to make the fetus visible; as a result of this, information about the health condition of the fetus becomes available, making it a 'potential patient'; against this background, in a society that allows abortion, ultrasound becomes a technology that could be used to prevent the birth of children with a specific health condition; and as a result of this, parents become decision-makers about the life of the fetus. Moral mediation, in other words, is a dynamic process of interpretation in which not only technological mediations but also human interpretations play a central role.

This hermeneutic lemniscate lays bare the dynamics of another far-reaching element of the moral significance of technology: the influence of technologies on *moral values and frameworks*. This phenomenon has been indicated as technomoral change' (Swierstra et al. 2009): technological developments affect morality itself. A good example here is the birth control pill, as analyzed by Annemarie Mol (1997). Because the pill disconnected sexuality from reproduction, it changed value frameworks regarding sexuality by normalizing sex that was not directed at reproduction. In doing so, it contributed to a growing acceptance of relations that cannot result in reproduction, like homosexual relations. Olya Kudina has shown that this phenomenon of value change can also be studied empirically, while it is in process. By studying empirically online discussions about Google Glass on YouTube, for instance, she has investigated how Glass has shifted the meaning of the concept of privacy, beyond the regular definitions found in textbooks (Kudina 2019; Kudina and Verbeek 2019). This phenomenon of technomoral change gives the ethics of technology an extra empirical-philosophical dimension: to evaluate technologies ethically, ethicists do not only need to anticipate their future social implications, but also the impact these technologies might have on the moral frameworks from which they might be evaluated in the future.

4.3 Morality in Design

As a result of this attention to the moral significance of technologies, the ethics of technology has also started to reach out more explicitly to design. When technologies are morally significant, after all, the ethics of technology can not only result in interesting *analyses* but also in better *technologies*. This focus on design ethics can be seen as a next step in deepening the engagement of philosophy of technology with actual technological artifacts and practices. One of the most influential approaches here has been Batya Friedman's value-sensitive design approach, which enables designers to anticipate the values at stake in technology design, in order to feed this back into the design process (Friedman and Hendry 2019; see also Van den Hoven et al. 2017).

The approach of moral mediation has been used to expand this program of value-sensitive design. On the one hand, it has been used to take value dynamism into account when designing technologies: rather than "loading" technologies with predefined values, design then becomes an intervention in the *dynamics* between humans, values, and technologies (Smits et al. 2019; Verbeek 2013; 2017). Values are not given, but develop in close interaction with the technologies we evaluate with the help of these very values. On the other hand, the moral design of technologies has been connected to political theories, aiming to arrive at a democratic moralization of technologies (cf. Verbeek 2020). A good example here is the work of Ching Hung, whose dissertation "Design for Green" (Hung 2019) investigates the ethical and political dimensions of behavior-guiding technologies in relation to environmental issues. By connecting mediation theory to behaviorism (Skinner 1971), libertarian paternalism (Sunstein and Thaler

2009), and agnostic democracy (Mouffe 2013), Hung develops a design approach that takes the political dimension of value sensitive design as a starting point.

5. Beyond the Empirical Turn

Where has the empirical turn brought the field? While much current work in philosophy of technology is explicitly 'empirically inspired,' there are also worries that the empirical turn gave up too much of the "classical" perspective. Within the phenomenological approach, this critique has been mainly voiced by North-American philosopher Robert Scharff (2012) and Dutch philosophers Jochem Zwier, Vincent Blok, and Pieter Lemmens. The empirical orientation of postphenomenology, they hold, results in a focus on the micro-level of material artifacts, while losing the macro-level of overarching interpretive frameworks and power structures out of sight (Zwier et al. 2016). While the transcendental approach of classical philosophy may have overlooked the concrete details of actual technologies, the empirical approach of contemporary philosophy risks to lose the bigger picture.

The work of some current thinkers in philosophy of technology indeed seems to move back toward a more transcendentally oriented approach, such as the work of North-American philosopher Nolen Gertz, who focuses on the intrinsically nihilistic character of contemporary technology (Gertz 2019). Still, these new approaches do not abandon the empirical-philosophical orientation. Rather, they use empirical studies as a basis for analyses at a more transcendental level. In fact, the philosophy of technology seems to be entering a new stage, with new connections between the empirical and the philosophical. I will elaborate two of the directions this development could take, one focusing on the new philosophical questions to ask in connection with actual technological developments (philosophy from technology), the other on new forms of philosophical engagement with technology and its social roles and implications (philosophy for technology).

5.1 Philosophy from Technology: Technological Mediation and Conceptual Disruption

The recent work done on technological mediation, as part of the postphenomenological approach in philosophy of technology, can be seen as one example of the new philosophical connections to empirical work. The approach of technological mediation investigates how human-technology relations can be included in philosophical subdisciplines such as the philosophy of science, ethics and political philosophy, and metaphysics (Verbeek 2016). Its central idea is that technologies help human beings to answer central philosophical questions, like the three questions that Immanuel Kant considered to be essential: What can I know? What should I do? What may I hope for?

Regarding knowledge, for instance, the central focus of the mediation approach is on the ways in which technologies help scientists to perceive and interpret the phenomena they study. Within the field of neuroscience, for example, technologies like fMRI play a crucial role in how scientists understand the brain and human functioning. As the work of Bas de Boer shows, these technologies are not just scientific instruments, but they also play a constitutive role in the interpretive frameworks of scientists: they help to shape how scientists understand phenomena such as 'visual attention' and the complexity of the brain (De Boer et al. 2018; De Boer et al. 2020).

Within ethical and political philosophy, a similar movement can be made. Not only do technologies play a mediating role in ethical actions, decisions, and frameworks, as the discussion on value dynamism in the previous section of this chapter made clear; their mediating role also has a political dimension (cf. Verbeek 2020). Technologies can mediate *power relations* and *political interpretations*, for instance, as Robert Rosenberger shows in his work on street furniture that discriminates against unhoused people (Rosenberger 2018). Technologies can also mediate *political interactions*, as exemplified by social media that sometimes lock people up in their filter bubbles. And they can contribute to the formation of *political issues:* the 'citizen sensing' movement, for instance, encourages citizens to use sensors to put things on the political agenda, for instance by detecting radiation, measuring the ground water level, or monitoring airplane noise near airports (Woods et al. 2018).

Also in the realm of metaphysics, the mediation approach can bring a new perspective. The relation between technology and religion can be a good starting point here. Just like science, technology is often opposed to religion: technological manipulation and intervention seem to be at odds with the religious openness for transcendence, for what *cannot* be controlled and manipulated. But in fact, people's encounter with this transcendence typically takes shape *via* technology (Aydin and Verbeek 2015). In vitro fertilization is not simply a technology to make a child, but it also reveals how un-makeable life is, for people who cannot get pregnant without this technology (Verbeek 2010). Also our understanding of death—the other boundary of life—is technologically mediated: neurotechnologies, for instance, have brought the new category of 'brain death' (De Boer and Hoek 2020). Rather than being opposed to it, technologies mediate what transcendence can mean for human beings.

In a sense, the approach of technological mediation can be seen as a 're-transcendentalization' of the philosophy of technology, via the empirical turn. While the empirical turn aimed to move away from the transcendentalist reduction of technologies to their conditions, its focus on actual human-technology relations has made visible that technology is in fact part of the *human* condition: the relations between human beings and the world are always conditioned by the technological medium in which these relations play themselves out. This position is a continuation of earlier philosophical-anthropological insights in the technological character of human existence, which has been analyzed with notions such as 'originary technicity' (Leroi-Gourhan 1993), 'originary prostheticity' (Stiegler 1998), 'essential deficiency' (Gehlen 1998), and 'natural artificiality' (Plessner 2019). The mediation approach investigates the consequences of this 'technological condition' for human thinking, and therefore for philosophy itself. In fact, IT turns the philosophy of technology into a 'philosophy *from* technology': it takes concrete human-technology relations as a starting point for re-thinking the basic questions of philosophy.

A comparable move is made in current work in ethical theory of technology. In 2019, a large consortium of Dutch researchers received funding for a 10-year research program (2019–2029) on the Ethics of Socially Disruptive Technologies, which can be seen as a radical philosophical consequence of the empirical turn.[1] The project starts from the observation that technologies can disrupt ethical concepts and categories. Technologies such as robots, gene editing, and climate engineering require a revision of the concepts with which ethical theory has been working. How do we use the concept of 'moral agency' if robotic technologies such as self-driving cars have 'learning' algorithms that enable them to make moral decisions about the lives of human beings when a crash occurs? If the DNA of an organism contains both human and nonhuman animal elements, should we consider this organism to have animal rights, human rights, or both? How shall we determine whether the risks connected to climate engineering technologies like 'dimming the sun' (Roeser et al. 2019) are acceptable: how to represent future generations and nature itself in democratic processes to decide about this, and how to use the concept of intrinsic value when nature has become an engineering project?

In all of these cases, concrete technologies and technological developments require the revision of ethical frameworks and the development of new concepts. To understand and evaluate technologies, we have to construct our conceptual frameworks while we are using them. Technologies are not merely 'objects' of philosophical reflection here, which can be studied with empirical-philosophical methods. Rather, they *challenge* philosophical and ethical theories, and reveal that the vocabularies, approaches, and concepts that have been developed over the past centuries need to be expanded, updated, and revised. The empirical turn, therefore, is not a turn away from philosophy, but a turn toward a new direction in philosophy.

5.2 Philosophy for Technology: Guidance Ethics

The empirical turn did not only have an impact on the philosophy of technology itself, but also on its relations with society. Besides bringing new connections between philosophy and technology, it also resulted in a new approach in applied ethics. The insight that technology plays a conditional and constitutive role in society and human existence has resulted in an alternative to the biomedical model of applied ethics. Medical ethics typically focuses on 'ethical assessment,' often executed by medical-ethical committees that evaluate proposals for research or intervention in order to approve or reject them. In the ethics of technology, though, the focus is rather on 'accompaniment.' Its relevance is not only to be found in the approval or rejection of technologies, but also in the guidance of their development, implementation, and use: precisely in this interplay between technology and society, values are at stake that need to be identified and taken into account in the practices around technologies.

The recently developed Guidance Ethics Approach (Verbeek and Tijink 2020; see Figure 3.1) is one manifestation of this new type of applied ethics. In this approach—which takes inspiration from the approach of citizen science (Vohland et al. 2021) and from positive design (Desmet and Pohlmeyer 2013)—ethical reflection is taken to the actual practices in which technologies are being used by citizens and professionals. In a three-step approach, it aims to (1) analyze the technology in its concrete context of use; (2) anticipate the potential implications of this technology for all relevant stakeholders, in order to identify the values that are at stake in these implications; and (3) translate these values into concrete action perspectives regarding the technology itself (redesign), its environment (regulation, reconfiguration), and its users (education, communication, empowerment).

Guidance ethics aims to be an ethics from within rather than from outside: it does not seek to find a distant position for technology assessment but rather a close connection to guide the technology in its trajectory through society. Also, it aims to do ethics bottom-up rather than top-down: instead of letting ethical experts apply existing ethical approaches to a technology, it invites professionals and citizens to voice the ethical concerns they encounter in their everyday dealing with the technology. And, third, guidance ethics is a form of 'positive ethics' rather than negative ethics. This does not

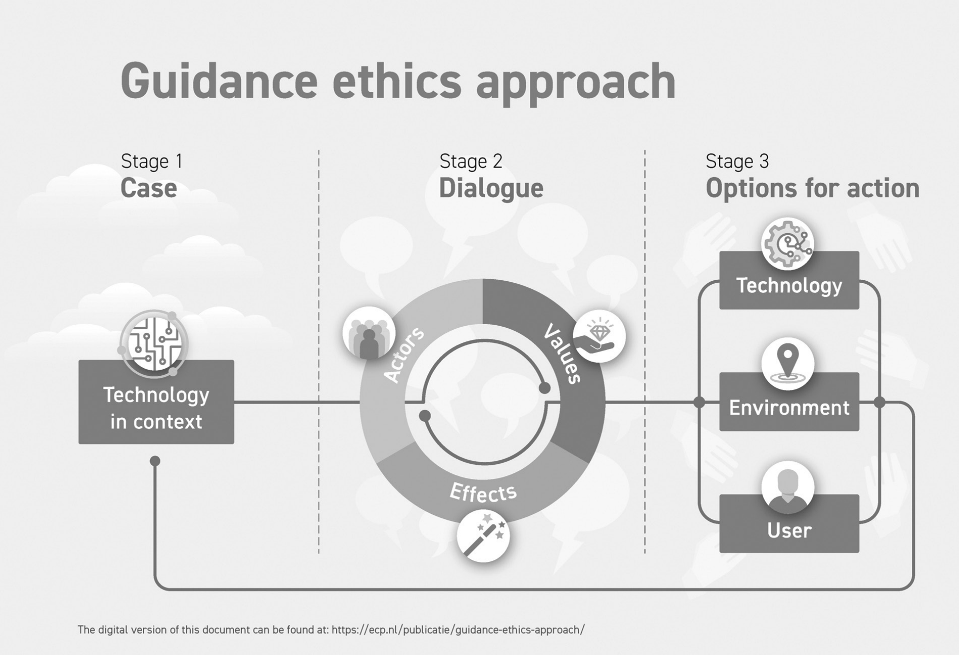

FIGURE 3.1: The Guidance Ethics Approach. © 2020 ECP | Platform voor de informatiesamenleving.

imply that the approach always has a positive evaluation of new technologies, but rather that its primary focus is not on defining the boundaries of what we do *not* want, but on identifying the conditions for what we *do* want. Along these lines, guidance ethics incorporates philosophical insights in the relations between technologies, human beings, and societies, and connects them to actual practices around technologies.

6. Conclusion

In the meantime, new directions are on the horizon already. A new generation of philosophers of technology, for instance, including Pak-Hang Wong, Tom Wang, and Ching Hung, is calling for a 'multicultural turn' in philosophy of technology. Given the global impact and implications of technology, they intend to expand debates in philosophy of technology by bringing in perspectives and approaches from outside the Western frameworks that are currently dominating the field (Wong and Wang 2021). Others are calling for a 'terrestrial turn,' in view of the environmental crisis (Lemmens et al. 2017). The notion of the 'Anthropocene,' indicating the current, human-dominated stage of development of planet Earth, inspires philosophers to thematize the "planetary condition" of humankind, and the role of technology in it. Rather than zooming in on concrete technologies, as the empirical turn proposed, we need to zoom out toward the planet and our technological way of dealing with it. At the same time, some call for a 'political turn' in the philosophy of technology, again focusing on larger societal and cultural patterns, power relations, and interpretative frameworks that need to be understood more closely in order to be able to engage with them politically (Gertz 2020).

There is no doubt that the philosophy of technology will still take many turns after the empirical turn it last made. What can be concluded for now is that the empirical turn has become a defining characteristic of the field. Taking its basis not only in the philosophical tradition but also in concrete engagement with actual technologies, it has expanded the scope of philosophy *of* technology toward philosophy *from* technology, when technological developments challenge existing philosophical frameworks, and philosophy *for* technology, when philosophical reflection is integrated in technological design, implementation and use. Rather than making the philosophy of technology less philosophical, as the oxymoron of "empirical philosophy" might suggest, the empirical turn has strengthened its philosophical rigor and ambition. It has laid the foundation for the unique ability of philosophy of technology to combine profound philosophical innovation with empirical and societal engagement, which, it is hoped, will serve as a strong basis for any future turns that the philosophy of technology might take.

Note

[1] For more information, see: https://www.esdit.nl.

References

Achterhuis, Hans, ed. 2001. *American Philosophy of Technology: The Empirical Turn*. Bloomington and Indianapolis: Indiana University Press.

Aydin, Ciano, and Peter-Paul Verbeek. 2015. "Transcendence in Technology." *Technè: Research in Philosophy and Technology* 19 (3): 291–313.

Borgmann, Albert. 1984. *Technology and the Character of Contemporary Life*. Chicago: University of Chicago Press.

Borgmann, Albert. 1995. "The Moral Significance of the Material Culture." In *Technology and the Politics of Knowledge*, edited by A. Feenberg and A. Hannay, 85–93. Bloomington: Indiana University Press.

Brey, Philip. 2010. "Philosophy of Technology after the Empirical Turn." *Techné: Research in Philosophy and Technology* 14 (1): 36–48.

De Boer, Bas, and Jonne Hoek. 2020. "The Advance of Technoscience and the Problem of Death Determination: A Promethean Puzzle." *Techné: Research in Philosophy and Technology* 24 (3): 306–331.

De Boer, Bas, Hedwig Te Molder, and Peter-Paul Verbeek. 2018. "The Perspective of the Instruments: Mediating Collectivity." *Foundations of Science* 23 (4): 739–755.

De Boer, Bas, Hedwig Te Molder, and Peter-Paul Verbeek. 2020. "Constituting 'Visual Attention': On the Mediating Role of Brain Stimulation and Brain Imaging Technologies in Neuroscientific Practice." *Science as Culture* 29 (4): 503–523.

Desmet, Pieter, and A. E. Pohlmeyer. 2013. "Positive Design: An Introduction to Design for Subjective Well-Being." *International Journal of Design* 7 (3): 5–19.

Ellul, Jacques. 1964. *The Technological Society*. New York: Alfred A. Knopf.

Feenberg, Andrew. 1991. *Critical Theory of Technology*. Oxford: Oxford University Press.

Feenberg, Andrew. 1999. *Questioning Technology*. London: Routledge.

Feenberg, Andrew. 2000. "From Essentialism to Constructivism: Philosophy of Technology at the Crossroads." In *Technology and the Good Life?* edited by E. Higgs, A. Light, and D. Strong, 294–315. Chicago: University of Chicago Press.

Franssen, Maarten, Pieter E. Vermaas, Peter Kroes, and Anthonie W. M. Meijers, eds. 2016. *Philosophy of Technology after the Empirical Turn*. Dordrecht: Springer.

Friedman, Batya, and D. Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press.

Gehlen, Arnold. 1998. *Man, His Nature and Place in the World*, trans. Clare McMillan and Karl Pillemer. New York: Columbia University Press.

Gertz, Nolen. 2019. *Nihilism and Technology*. Boston: MIT Press.

Gertz, Nolen. 2020. "Democratic Potentialities and Toxic Actualities: Feenberg, Ihde, Arendt, and the Internet." *Techné: Research in Philosophy and Technology* 24 (1/2): 178–194.

Heidegger, Martin. 1977. *The Question Concerning Technology*. In *The Question Concerning Technology and Other Essays*, trans. W. Lovett. New York: Harper and Row.

Hung, Ching. 2019. *Design for Green: Ethics and Politics for Behavior-steering Technology*. Enschede: University of Twente.

Ihde, Don. 1990. *Technology and the Lifeworld*. Bloomington: Indiana University Press.

Ihde, Don. 1993. *Postphenomenology*. Evanston: Northwestern University.

Illies, Christian, and Anthonie Meijers. 2009. "Artefacts without Agency." *Monist* 92 (3): 420–440.

Joerges, Bernward. 1999. "Do Politics Have Artefacts?" *Social Studies of Science* 29 (3): 411–431.

Kroes, Peter, and Anthonie Meijers, eds. 2001. *The Empirical Turn in the Philosophy of Technology*. London: JAI.

Kroes, Peter, and Anthonie Meijers. 2006. "The Dual Nature of Technical Artefacts." *Studies in History and Philosophy of Science* Part A 37 (1): 1–4.

Kroes, Peter, and Peter-Paul Verbeek, eds. 2014. *The Moral Status of Technical Artefacts*. Dordrecht: Springer.

Kudina, Olya. 2019. *The Technological Mediation of Morality: Value Dynamism and the Complex Interaction between Ethics and Technology*. Enschede: University of Twente.

Kudina, Olya, and Peter-Paul Verbeek. 2019. "Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy." *Science, Technology, and Human Values* 44 (2): 291–314.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1993. *We Have Never Been Modern*, trans. C. Porter. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1999. *Pandora's Hope*. Cambridge, MA: Harvard University Press.

Lemmens, Pieter, Vincent Blok, and Jochem Zwier. 2017. "Toward a Terrestrial Turn in Philosophy of Technology." *Techne: Research in Philosophy and Technology* 21 (2–3): 114–126.

Leroi-Gourhan, André. 1993. *Gesture and Speech*. Boston: MIT Press. [*Le Geste et La Parole: tome 1, Technique et Langage*. 1964, tome 2, La mémoire et les rythmes. Paris: Albin Michel, 1965.]

Mol, Annemarie. 1997. "Wat is kiezen? Een empirisch-filosofische verkenning." Enschede: Universiteit Twente (inaugural lecture).

Mouffe, Chantal. 2013. *Agonistics: Thinking the World Politically*. London: Verso.

Peterson, Martin. 2012. "Three Objections to Verbeek." In *Book Symposium on Peter Paul Verbeek's Moralizing Technology: Understanding and Designing the Morality of Things, Philosophy and Technology*, edited by E. Selinger et al. *Philosophy and Technology* 25: 619–625. https://link.springer.com/article/10.1007/s13347-011-0058-z

Peterson, Martin, and Andreas Spahn. 2010. "Can Technological Artefacts Be Moral Agents?" *Science and Engineering Ethics* 17 (3): 411–424.

Pitt, Joseph C. 2014. " 'Guns Don't Kill, People Kill': Values in and/or around Technologies." In *The Moral Status of Technical Artefacts*, edited by P. Kroes and P.-P. Verbeek, 89–101. Dordrecht: Springer.

Plessner, Helmuth. 2019. *Levels of Organic Life and the Human: An Introduction to Philosophical Anthropology*, trans. Millay Hyatt. New York: Fordham University Press.

Roeser, Sabine, Behnam Taebi, and Neelke Doorn. 2019. "Geoengineering the Climate and Ethical Challenges: What We Can Learn from Moral Emotions and Art." *Critical Review of International Social and Political Philosophy* 23 (5): 641–658.

Rosenberger, Robert. 2018. *Callous Objects: Designs against the Homeless*. Minneapolis: University of Minnesota Press.

Rosenberger, Robert, and Peter-Paul Verbeek. 2015. *Postphenomenological Investigations: Essays on Human-Technology Relations*. Lanham: Lexington.

Scharff, Robert C. 2012. "Empirical Technoscience Studies in a Comtean World: Too Much Concreteness?" *Philosophy and Technology* 25: 153–177.

Selinger, Evan. 2006. *Postphenomenology: A Critical Companion to Ihde*. New York: SUNY Press.

Skinner, Burrhus Frederic. 1971. *Beyond Freedom and Dignity*. Middlesex, England: Penguin Books.

Smits, Merlijn, Bas Bredie, Harry Van Goor, and Peter-Paul Verbeek. 2019. "Values that Matter: Mediation Theory and Design for Values." Academy for Design Innovation Management Conference 2019: Research Perspectives in the Era of Transformations, 396–407. London: Loughborough University.

Stiegler, Bernard. 1998. *Technics and Time I: The Fault of Epimetheus*. Stanford: Stanford University Press.

Sunstein, Cass, and Richard Thaler. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*, revised & expanded edition. New York: Penguin Books.

Swierstra, Tsjalling, Dirk Stemerding, and Marianne Boenink. 2009. "Exploring Techno-moral Change: The Case of the ObesityPill." In *Evaluating New Technologies*, edited by P. Sollie and M. Duwell, 119–138. Dordrecht: Springer.

Van den Hoven, Jeroen, Seumas Miller, and Thomas Pogge, eds. 2017. *Designing in Ethics*. Cambridge: Cambridge University Press.

Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Penn State University Press.

Verbeek, Peter-Paul. 2008. "Obstetric Ultrasound and the Technological Mediation of Morality: A Postphenomenological Analysis." *Human Studies* 31: 11–26.

Verbeek, Peter-Paul. 2010. "Designing the Human Condition: Reflections on Technology and Religion." *ET Studies* 1 (1): 39–52.

Verbeek, Peter-Paul. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press.

Verbeek, Peter-Paul. 2013. "Technology Design as Experimental Ethics." In *Ethics on the Laboratory Floor*, edited by S. van den Burg and Tsj. Swierstra, 83–100. Basingstoke: Palgrave Macmillan.

Verbeek, Peter-Paul. 2014. "Some Misunderstandings About the Moral Significance of Technology." In *The Moral Status of Technical Artefacts*, edited by P. Kroes and P.P. Verbeek, 75–88. Dordrecht: Springer Netherlands.

Verbeek, Peter-Paul. 2015. "Beyond Interaction: A Short Introduction to Mediation Theory." *Interactions* 22 (3): 26–31. ISSN 1072-5520.

Verbeek, Peter-Paul. 2016. "Toward a Theory of Technological Mediation: A Program for Postphenomenological Research." In Technoscience and Postphenomenology: The Manhattan Papers, edited by J. K. Berg, O. Friis, and Robert C. Crease, 189–204. London: Lexington Books.

Verbeek, Peter-Paul. 2017. "Designing the Morality of Things: The Ethics of Behavior-Guiding Technology." In *Designing in Ethics*, edited by J. van den Hoven, S. Miller, and Th. Pogge, 78–94. Cambridge: Cambridge University Press.

Verbeek, Peter-Paul. 2020. "Politicizing Postphenomenology." In *Reimagining Philosophy and Technology, Reinventing Ihde*, edited by G. Miller and A. Shew, 141–155. Cham: Springer.

Verbeek, Peter-Paul, and Daniel Tijink. 2020. *Guidance Ethics Approach*. The Hague: ECP.

Vohland, Katrin, Anne Land, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Reoland Samson, and Katherin Wagenknecht, eds. 2021. *The Science of Citizen Science*. Cham: Springer.

Winner, Langdon. 1986. "Do Artifacts Have Politics?" In *The Whale and the Reactor*. Chicago: University of Chicago Press.

Wong, Pak H., and T. X. Wang, eds. 2021. *Harmonious Technology: A Confucian Ethics of Technology*. London: Routledge.

Woods, Melanie, Mara Balestrini, Sihana Bejtullahu, Stefano Bocconi, Gijs Boerwinkel, Marc Boonstra, Douwe-Sjoerd Boschman, Guillem Camprodo, Saskia Coulson, Tomas Diez, Ioan Fazey, Drew Hemment, Christine van den Horn, Trim Illazi, Ivonne Jansen Dings, Frank Kresin, Dan MacQuillan, Susana Nascimento, Emma Pareschi, Alexandre Polvora, Ron Salaj, Michelle Scott, and Gui Seiz. 2018. *Citizen Sensing: A Toolkit*. Making Sense. https://discovery.dundee.ac.uk/files/39984081/Citizen_Sensing_A_Toolkit.pdf.

Woolgar, Steve, and G. Cooper. 1999. "Do Artefacts Have Ambivalence?" *Social Studies of Science* 29 (3): 433–449.

Zwier, Jochem, Vincent Blok, and Pieter Lemmens. 2016. "Phenomenology and the Empirical Turn: A Phenomenological Analysis of Postphenomenology." *Philosophy and Technology* 29 (4): 313–333.

CHAPTER 4

PHILOSOPHY OF TECHNOLOGY AND THE CONTINENTAL AND ANALYTIC TRADITIONS

MAARTEN FRANSSEN

1. Introduction

Philosophy of technology is a young field, one of the youngest among “philosophies of,” but since in philosophy development is measured in millennia, this can be a misleading statement. As an academic discipline, philosophy of technology has now had half a century to prove itself. The results are mixed. At the turn of the millennium Peter Kroes and Anthonie Meijers (2000), in their introduction to *The Empirical Turn in the Philosophy of Technology*, described the field as “a discipline in search of its identity.” More recently Franssen and Koller (2016) claimed that this situation had not improved in the fifteen years since and that the field is still lacking in substantive unity and systematicity. It is of crucial importance for the further development of philosophy of technology to arrive at an understanding of what underlies this situation—even if one disagrees with the assessment, it is still significant that this is how the field’s condition is judged by some of its practitioners—and what can be done about it.

The formative years of philosophy of technology are, roughly, the two decades between 1965 and 1985. This is not to say that before that, there was no philosophy of technology at all. The first book to explicitly offer a “philosophical of technology,” Ernst Kapp’s *Grundlinien einer Philosophie der Technik*, dates from 1877, and it was followed by other books with a (seeming) reference to philosophy and technology in their title, by Zschimmer (1914), Dessauer (1927), and Schröter (1934) in German and by Engelmeier (1912) in Russian. After the Second World War, books continued to be published, again mostly in German, which “addressed” technology, for example by Jünger (1949) and Gehlen (1957). But none of these books was instrumental in establishing a philosophical field. The views expressed were often highly idiosyncratic, particularly so in the case of Kapp and Dessauer, and even the more recent ones were not written as contributions to a recognized field, let alone an academic discipline, or were in a systematic way grounded in philosophy. Philosophy of technology as an academic discipline I see as emerging with the symposium that the journal *Technology and Culture* organized in 1965, with contributions by Mumford, Feibleman, Skolimowski, Jarvie, and Bunge. Several of these papers were reprinted in the first anthology of texts that presented a wide variety of views and orientations, Carl Mitcham's and Robert Mackey's *Readings in the Philosophical Problems of Technology* (1972). With this book, the academic discipline of philosophy of technology was finally in existence.

As expected, the authors of these founding publications showed a keen awareness that a characterization of the field must be attempted, and they did so mainly by listing and classifying the questions they saw as central to the field and the problems it sets out to address. Melvin Kranzberg, editor-in-chief of *Technology and Culture*, mentioned "the questioning of technology in terms of human values"; "the attempt to define technology"; "the epistemological analysis of technology"; and "the investigation of the rationale for technological development" (1966). Mitcham and Mackey (1972) distinguish two classes of questions: on the one hand, inquiries into the logic and epistemology of technology and, on the other hand, the meaning of technology, primarily the ethical and political meaning. In his *Philosophy of Technology*, the first introductory textbook for the field, Frederick Ferré spent thirteen pages on a systematic attempt to define technology (1988, 14–26).

Acknowledgment of the relevance of this effort, however, seems to have declined rapidly in later work. Don Ihde's first book, *Technics and Praxis* (1979), contains no attempt to position it with respect to the major questions of a field that was then still *in statu nascendi*. It right away focuses on internal questions. In his *Philosophy of Technology: An Introduction* (1993), Ihde hardly spent a page on the meaning of technology and merely listed the three "components" that a definition would have to include (Ihde 1993a, 47). A similar attitude of bypassing definitional issues can be seen in Peter-Paul Verbeek's more recent books, which, put very briefly, address the questions how we can do justice to the roles that material objects play in our lives (Verbeek 2005, 1–3)—where it is assumed that currently no justice is done to these roles– and how we can do justice to the moral dimensions of material objects (Verbeek 2011, 2)—where it is assumed that material objects have these "dimensions." But whereas these authors appear simply to have taken for granted that they were contributing to a field with a well-established identity and focus—even if this assumption is highly questionable –, Kroes and Meijers (2000), in their assessment in which they precisely questioned this assumption, did not put much effort either into articulating the new questions for the philosophy of technology for which they claimed their empirical turn would make room. They did so only for the very specific ones that underlay their personal research project on "the dual nature of technical artifacts." Nor did they clarify to what extent these new questions should be seen as replacing older questions or as complementary to them.

That neither Ihde's *Philosophy of Technology: An Introduction* nor Ferré's *Philosophy of Technology* were reissued after the 1995 second edition of the latter, and that no introductory textbook of a similar scope and with a similar general audience in mind has been published since Joseph Pitt's *Thinking about Technology* (1999), I take to be symptomatic of the situation. Consider, for comparison, the situation in the philosophy of science, where since 1998 eleven new introductory textbooks were published. One gets the impression that already, very soon after its birth, the field of philosophy of technology appeared to its practitioners as too extensive and too variegated to be fully graspable.

What may have contributed to this situation is that exactly during the formative decades of philosophy of technology, its philosophical environment was subject to significant reorientations. Two developments stand out. The first was that in this period the adjacent field of the philosophy of science entered a state of turmoil. The publication of Kuhn's *The Structure of Scientific Revolutions* in 1962 initiated a debate that, unlike previous debates in the philosophy of science, such as the controversy between Popperians and logical empiricists on the status of inductive reasoning, quickly spread to the social sciences and the humanities and caused a massive interest in the philosophy of science which lasted for several decades. This debate was dominated by views that were distinctly critical of the "received view" of science and of the support that this view had received from "traditional" philosophy of science. The second development was that the English-speaking academic world was opening up fast to new ideas from another, somewhat exotic world. In 1965 Heidegger was still a marginal figure on the horizon, of whom only *Being and Time* (1962, English translation) and *An Introduction to Metaphysics* were then available in English translation, plus a few essays. But from 1966 over a twelve-year period the entire corpus of his post-war publications was published, mostly by Harper and Row. The works of other philosophers quickly followed, and the term "continental philosophy" came in use to refer to the type of philosophy represented by these works.[1]

Whatever the causal factors that interfered, and in whatever way they interfered—and this is not the place to undertake a detailed historical investigation—it is difficult to see how a philosophical field can sustain itself without a shared understanding of the meaning of the term that defines it, or minimally, a shared understanding of the main controversies surrounding the meaning of that term. It is bound to fragment into small clusters of authors who understand one another's work because they happen to use key terms in the same sense. As I argue in this chapter, philosophers of technology fail to appreciate the extent to which different authors build upon different conceptions of technology and the extent to which this makes their claims and views and assessments incomparable. And if comparison is imposed nevertheless, what results is a cacophony, not a discipline.

In the next section I first address the question whether philosophy of technology's lack of unity can be seen to be (partly) due to its being "contested" between analytic and continental philosophy. In the subsequent section I address how the field's apparent failure to be aware that it lacks a shared conception of technology mars its development, and that as a "phenomenon" technology may even escape all attempts to define it. I then offer a response to this diagnosis by proposing to look upon the philosophy of technology as consisting of three subfields, and by making some suggestions concerning how the strengths of analytic and continental philosophy, as each sees them, can be brought to bear on these subfields.

2. Is Philosophy of Technology Balancing Astride the Analytic-Continental Divide?

From the 1980s on, those who arrived fresh at the field would be excused for thinking not only that there is a continental variety of philosophy but even that philosophy of technology originated as a continental discipline. In his first book, *Technics and Praxis*, Ihde described Heidegger as "among the first to raise technology to a central concern for philosophy" and stated that "Heidegger's philosophy of technology is one of the most penetrating to date" (1979, 103). In his later *Postphenomenology,* Ihde still described Heidegger as "surely one of the most important founders of the philosophy of technology" (1993b, 1).[2] Scholars of the "second generation," particularly Verbeek, have followed Ihde in this: Verbeek classifies Heidegger as the prime representative of "classical philosophy of technology." Scholars who wholeheartedly reject Heidegger's views nevertheless concur with this view; Andrew Feenberg, for example, pronounced Heidegger to be "no doubt the most influential philosopher of technology in this century" (1999, 183).

The contrast between analytic and continental philosophy as two "schools" or "cultures" of philosophy has been at play throughout the discipline for almost half a century. And with Heidegger termed a key figure, philosophy of technology is implicitly positioned in close proximity to continental philosophy. But in order to investigate whether there is any substance to the suggestion that philosophy of technology is contested between two major philosophical schools or cultures, analytic and continental, we must start with characterizing them. That, however, is far from easy. Both terms, it seems to me, refer jointly, and vaguely at that, both to concrete positions adopted in the course of the historical development of philosophy and much more diffuse attitudes or orientations that current philosophers assume which are felt to derive from this historical position-taking. So far only analytic philosophy has itself become the object of historical-philosophical study.[3]

Considered typical for analytic philosophy are an emphasis on the concepts and terms of philosophical discourse and questioning, a focus on analysis and cutting up large questions and problems into smaller bits, and a striving for clarity and rigor, often through the application of formal logic. This applies in particular to the heydays of "core" analytic philosophy, roughly from the late 1920s to the late 1950s. Since then, analytic philosophy as a label refers more to a method or style of doing philosophy, which looks at “core” analytic philosophy for its inspiration. This description serves ill to distinguish analytic philosophy from its opponent, continental philosophy, as if continental philosophy would explicitly reject analysis, clarity, and rigor. How could any form of philosophy get away with that? At most one can say that these are not emphasized or singled out as characterizing continental philosophy. But what does characterize continental philosophy is even harder to say than analytic philosophy, and any attempt to do so is bound to meet even more controversy. For current continental philosophy it is probably even more true that it is foremost a method and a style of doing philosophy, shaped, just like analytic philosophy, by a historical core. For the purpose of this chapter, I settle for an attempt to sketch and contrast these two cores.

Whatever the differences, the two philosophies share a radical origin. Both core analytical philosophy and core continental philosophy looked upon themselves as rejecting philosophy as it had been done before and as making a new beginning, if not revolutionizing philosophy, and both did so at the same time, during the 1920s. The character of their new beginning is of course what sets them apart. The distinction most relevant to our discussion is how they both positioned themselves with respect to science. Science formed the great challenge to 19th-century philosophy. The philosophical inquiry into the structure of the world and the attempt to make sense of it all had resulted, from the 17th century on, in the special sciences separating from philosophy one by one, starting with physics and ending with psychology in the late 19th century. This led to the question what was left for philosophy to do, if anything, and how it should go about doing it in the light of the totality of science, which continued to develop according to its own dynamic. The range of options for philosophy was vast, from Wittgenstein’s emaciated therapeutic conception to Husserl’s substantial ground-laying conception. But within that range, the cores of (what would become) analytic and continental philosophy emerged around two radically different positions with respect to science: analytic philosophy as accepting science as a background for philosophy, and continental philosophy as rejecting it. The analytic philosophy saw in formal logic a language that it shared with science. One of the challenges put to philosophy by science was that it threatened to leave no place for meaning. Analytic philosophy’s acceptance of science as a background entails that only so much of meaningfulness can be secured as science allows for. Continental philosophy rejected this approach as not delivering sufficient meaningfulness, or the right sort of meaningfulness, and therefore rejected science as a background, or rather saw little reason to engage with science at all.

Notwithstanding its original revolutionary zeal, analytic philosophy stopped being a revolutionary movement long before the formative period of philosophy of technology. It gradually became the new establishment, in the sense that most academic philosophy in the English-speaking world is done by philosophers who, by training rather than by well-considered choice, work in the analytic tradition. In the course of this development, it spread to include all the philosophical fields that analytic philosophy originally had wanted to kick out, in particular metaphysics and ethics. This allowed continental philosophy, which developed more slowly and less linearly, to prolong its status as revolutionary philosophy in opposition to the analytic establishment in the English-speaking world in the 1970s to 1990s.

One difference between analytic and continental philosophy is of particular importance here. Whereas core analytic philosophy involved philosophers in at least three countries—Austria, Germany, and England—continental philosophy I see as dominated by the single figure of Martin Heidegger. Though continental philosophy has "father figures"—Dilthey, Husserl—in the same way that analytic philosophy has "father figures"—Russell, Moore, Frege—who were a major source of inspiration but are generally seen as not themselves belonging to its core, core continental philosophy to me is the philosophy which developed in the space opened up by Heidegger's particular breach with traditional academic philosophy.[4] To investigate Heidegger's alleged role as one of the founders of philosophy of technology *is* therefore to investigate to what extent there is a continental philosophy of technology. This alleged role of Heidegger is based on two concrete elements of his work. The first is section 15 of his *Being and Time*, the second his essay "The Question Concerning Technology." These two are wide apart, however.

In section 15 of *Being and Time*, Heidegger describes how *Dasein*—his term for the conscious, living human being—is from the very start experiencing itself as living not so much in a world but rather *the* world, and that this world first of all has the character of a network of things that are meaningful in that they are "for something." A hammer is first of all something used "self-evidently" for hammering, and a pen for writing, and these activities take place in a house which is self-evidently for living in, and so forth. It requires an effort on the side of *Dasein*, which may be occasioned by the thing itself, when, for example, it is broken or ill fit to the task—when a hammer is too heavy, for instance—to conceive of a thing as an object, something that stands alone, independent of us or the other things in the world, with properties—such as size, weight, and material composition—that belong to it independently of us or other things. Heidegger distinguishes these two modes of being by calling the first, their being "equipment" and "for something," ready-to-hand (*Zuhanden*), and the second, their being objects independently of any meaningful context, present-at-hand (*Vorhanden*).

To what extent does this description belong to philosophy of technology, or can even be the first step toward a philosophy of technology that it is so regularly portrayed to be? Heidegger's description does not require the ready-to-hand things that make up our world to be artifacts. Although the examples that Heidegger himself mentions—hammers, pens, houses—are all artifacts, the neutral term "equipment" (*Zeug*) that he uses for things in their ready-to-hand mode seems to me intentionally neutral in this respect. Heidegger seems not interested at all in the origins of equipment as artifacts. The word *Technik* occurs exactly once in *Being and Time*, in a parenthetical clause in section 61. The distinction between ready-to-hand and present-at-hand, of which so much is made, is on further reflection quite superficial.[5] Perhaps it takes the perspective of a child rather than the no-nonsense perspective of an adult to bring this out. Imagine how a child will take in the world in which it lives. First of all, it consists of natural objects just as much as artificial objects, and among both some are ready-to-hand but others not.[6] Stones are for throwing at birds or skimming on the water, leaves of grass are for being stretched between your fingers and being blown to produce their shrill note, feathers are for being put in trophy boxes, and so forth. But many other things are just there: broken-off fragments of wood, unidentifiable components of equipment long gone, and the like. The world delivers us with an entire spectrum of ready-to-hand and present-at-hand manifestations.

Specifically from the perspective of technology and philosophical reflection on technology, Heidegger's picture in *Being and Time* is highly unsatisfactory. Not only does it seem immaterial to Heidegger how equipment has come to make up the world and how this network of meaningful relations has been formed; his description of use is simply inadequate. When hammering in a nail with a hammer, the nail is just as important as the hammer. Even if it is granted that one need not consider the weight of the hammer before taking it up, and therefore that one need never conceive of the hammer as having a certain weight prior to grasping it, this cannot be so for nails. If one who is engaged in a carpentry job does not consider how many nails are needed, how long and how thick they should be, and at what distances they should be placed, one is not likely to end up with something that remains standing. According to Heidegger, however, this taking of things to have inherent properties is exactly what comes only when things are looked upon as present-to-hand, a development out of the ready-to-hand, and also, in Heidegger's view, the first fateful step in the direction of metaphysics. It is difficult to see how the calculative-planning thought so scorned by Heidegger can be absent even when life proceeds in the taken-for-granted way that Heidegger identifies with the ready-to-hand. But then the details of these relations seem not to be at all what he is interested in. In *Being and Time,* Heidegger is exclusively engaged in an analysis of *Dasein*, which to him is equivalent to living-in-the-world—not an analysis of the world in the traditional philosophical sense of "what there is," "reality." That sort of analysis, as belonging to metaphysics, is precisely what Heidegger intends to destroy, as he announces in the introduction to the book.

In Heidegger's essay "The question concerning technology" (1977), in contrast to *Being and Time*, the word *Technik* takes center stage. But it is infused with the calculative thinking which goes together with the mode of being that is present-at-hand. *Technik* is exactly *not* what underlies the ready-to-hand. Heidegger does not define technology as the making and using of artifacts, in order to distinguish next between "traditional" technology and modern technology. Such a definition of technology he terms "untenable." Constructions like machines and power stations, as well as the type of human action that is the making and the using of such things, are to Heidegger rather symptoms or manifestations of something much larger, and it is this much larger "something" that he calls *Technik*. It comes with a "a way of revealing," a bringing into the open (a Heideggerian technical term) which, in contrast to artistic or craftsmanly creation—an individual act directed at the creation of a single concrete thing—is a constant challenging of nature to be available for nothing but further challenging, in an endless and ever-expanding cycling and recycling. *Technik* is that of which this "way of revealing," which Heidegger calls das *Ge-stell* (translated as "Enframing") "is" the essence, or that in the essence of which das *Ge-stell* "holds sway," or "shows itself," or "lies." *Technik*, the totality underlying the "way of revealing" and its instruments—the things that are, in his view, misleadingly, referred to as "modern technology"—therefore manifests itself as much wider, way beyond these "symptoms." *Technik* "includes all the areas of beings which equip the whole of beings: objectified nature, the business of culture, manufactured politics, and the gloss of ideals overlying everything" (1973, 93). To Heidegger the term *Technik* "coincides with the term 'completed metaphysics'" (1973, 93)—a metaphysics that started with Plato and that ever since has proceeded, step by step, inexorably, in the direction taken.

This bodes ill for any suggestion that things could be otherwise. And indeed, although the "revealing" is a manifestation of human activity, *Technik*, as Heidegger emphasizes, is not. *Technik* and its *Ge-stell* result from a *Geschick* or "destining"; "mankind" has been "claimed" (*in Anspruch genommen*) into responding with the way of revealing that belongs to technology. As a consequence, and this is also emphasized by Heidegger, no form or amount of human action can "put an end" to the manifestations of modern technology, might we want to do so. We will simply have to wait until another destining will "claim" us to approach the cosmos, or "what there is," in another way.

To present Heidegger as a philosopher of technology, then, or even as pointing the way to a philosophy of technology, seems untenable. The two conceptions of technology that have been "constructed" from his work to justify this are in stark opposition to one another. Heidegger himself flatly rejected the idea that his questioning can be taken as a philosophy of technology. In his view, any philosophy of technology—any of the forms that philosophical reflection on technology has taken in the past century—cannot but be a pointless exercise within completed metaphysics. In the current crave for identifying almost anything as "post-," Heidegger's view with respect to philosophy deserves this prefix most: it is postphilosophy.

Indeed, if one looks into the details of how the work of philosophers of technology who advocate Heidegger as one of the principal philosophers of technology is in fact connected to Heidegger's views—or to any part of continental philosophy, for that matter—one discovers that these connections tend to be wafer-thin, if discernible at all. It is doubtful to what extent Ihde's characterization of his philosophy as grounded in phenomenology and as being itself "postphenomenology" can be taken seriously. There is no common ground between Heidegger's talk of the essence of technology—*das Wesen der Technik*—and the idea that the meaning of artifacts is "multi-stable" in Ihde's terminology. Similar problems arise for Verbeek, who more than any other current philosopher of technology seems to advocate a continental approach to the philosophy of technology. When he gives as his "elementary definition" of phenomenology that it is "the philosophical analysis of the structure of the relations between human beings and their lifeworld" (2011, 7) and describes as the "central phenomenological idea" that "human-world relations need to be understood in terms of 'intentionality'" (2011, 15), Husserl, who introduced the term phenomenology to modern philosophy, seems to be far away.[7] Verbeek, following Ihde, places intentionality in the relation between humans and their world (116), but for Husserl intentionality is entirely a phenomenon within human consciousness. The meaning of the term phenomenology, as understood by Husserl, and the extent to which Heidegger's philosophy is phenomenological in either Husserl's sense or any explicable sense, are notoriously difficult and contested issues. Heidegger himself was extremely condescending about Husserlian phenomenology in private and the term disappears completely from his work once he had succeeded in being nominated as Husserl's successor in Freiburg. When Husserl finally set himself to actually read *Being and Time* he likewise came to the conclusion that the work had nothing to do with what to him was phenomenology.

We should therefore be very hesitant to characterize the work of philosophers like Ihde and Verbeek as a continental approach to philosophy of technology, and to accept the existence of any systematic approach to the philosophy of technology that can be placed in the tradition of continental philosophy. This is not to deny that there have been significant contributions to the philosophy of technology from continental Europe, for example by Jacques Ellul (1954) and Gilbert Simondon (1958). Ellul's writings exercised a strong influence in the 1960s, at the time attempts started to organize the philosophy of technology into a field of research. In 1962 the journal *Technology and Culture* published the proceedings of an international conference that took its title of "The technical order" from Ellul's keynote address. Since then, however, Ellul's work has slowly drifted to the margin. Simondon's work, in contrast, has only recently been gaining interest. Still, neither of them can be connected to any particular philosophical tradition, or even to a tradition of what doing philosophy is in the first place. In Ellul's extensive list of references only a handful of philosophers, from either tradition, occur—Jaspers, Marcel, Ortega y Gasset, Russell. The only philosopher to figure in Simondon's much shorter list is Canguilhem.

A much stronger influence, especially on the work of Verbeek, has been exercised not by some philosophical view but by the field of inquiry called Science and Technology Studies, especially the theoretical approach known as constructivism. This is an approach to the study of, initially, science, but later extended to technology, which originated in the 1970s out of dissatisfaction with the way that science was studied by philosophers of science. Due to what was perceived as analytic philosophy's reverence for science, philosophy of science was taken to be satisfied with mere rational reconstruction of the success stories of science. Proponents of the "Strong program in the sociology of knowledge" sought to replace this with an approach in which science would be studied as a social phenomenon, by the empirical human sciences, as it was their task to study all social phenomena. Philosophy was distrusted for its insistence on a priori judgements.[8] Like both analytic philosophy and continental philosophy, constructivism aimed to be revolutionary, but like latter-day continental philosophy the establishment which it targeted was analytic in outlook. And like continental philosophy as well, it aimed to restore the primacy of the humanities in studying all human activity. Its conception of the humanities, however, excluded "philosophizing" and implied a solid naturalism.[9] This is precisely what philosophers who might agree to the primacy of the humanities find objectionable in it (e.g. Winner 1993).

If constructivism is often included in "the continental view," this may be largely due to it having been radicalized by the work of Latour and Callon, both Frenchmen, in the 1980s. However, this radicalization concerned the choice for a particular methodology—at most an issue in the philosophy of the human sciences, therefore, not philosophy *tout court*. The strong program had been weak on methodology—which particular models and theories from the human sciences should be used for its explanatory aims. Latour and Callon gave it a strong but at the same time more extreme methodological orientation—that of semiotics. And part of the "social studies" community flatly rejected this reorientation (Amsterdamska 1990, Bloor 1999). It was not, however, a reorientation in the direction of philosophy. Although in the 1960s and 1970s—the heydays of structuralism—semiotics and philosophy had a love affair, which did leave its mark on continental philosophy, Latourian constructivism remains true to the principles of the "social studies" approach: it is naturalistic and distrustful of philosophy. In Latour's major early works (1987, 1988), from the continental tradition only Deleuze, Foucault and Serres are referenced, and as other philosophers Fleck, Kuhn and Canguilhem. Greimas and Courtés's *Sémiotique: Dictionnaire raisonné de la théorie du langage* can be found referenced in almost anything Latour writes.

Indeed, Latourian constructivism seems to be a major source of inspiration for one of the most controversial themes introduced in current philosophy of technology by Verbeek, a theme that seems to have relieved the ghost of Heidegger as what makes any analytic philosopher hesitant to enter the field, namely the treatment of artifacts as agents (see Peterson 2012 and Verbeek's 2012 reply for a taste of the controversy).

We may conclude, then, that what *prima facie* seems, or is portrayed as, continental philosophy of technology is in fact a highly superficial and eclectic borrowing. Little attention is paid to where various adopted views came from, what they originally were meant to do and whether they are compatible at all and therefore can be mixed. There is a way of doing philosophy of technology that incorporates work from continental philosophy, but it would be misleading to refer to it as a, or even the, continental-philosophical way of doing philosophy of technology, because it is not continental philosophy.

Neither, however, can one say that there is such a thing as analytic philosophy of technology, though for different reasons.[10] When philosophy of technology began to take shape, core analytic philosophy was already over and few philosophers still saw themselves as representing it. To contrast analytic and continental philosophy was significant only when and where continental philosophy was present to a significant degree. Friedrich Rapp's *Analytical Philosophy of Technology* (1981) is not analytic philosophy: Rapp's use of "analytical philosophy" refers not to conceptual analysis but to empirical analysis. Somewhat different from the plea for an empirical turn made by Kroes and Meijers (2000) two decades later, Rapp urged philosophers to become much better informed about the historical development of technology before advocating "metaphysical" views concerning its degree of inevitability and its appreciation. Rapp takes issue only with the quality of extant metaphysical views, not their philosophical legitimacy. Then what *prima facie* seems, or is sometimes portrayed as, analytic philosophy of technology is rather philosophers exercising the only sort of philosophy they understand to be philosophy, simply because it is the philosophy they were educated in, and not of philosophers choosing and then implementing a particular approach to philosophy of technology from among several possible ones.

To look upon philosophy of technology as contested between the two approaches or continental and analytic philosophy, then, is not fruitful. Which is not to say that the distinction lacks all relevance. But to see how, we must first return to what I mentioned in the Introduction as being responsible for philosophy of technology's lack of identity: the field's failure to arrive at a shared understanding of the term "technology."

3. The All Too Many Meanings of "Technology"

Part of the argument developed in the previous section is that it is simply an error to assume that, when Heidegger is making claims about something he calls *Technik*, he is referring to that which the English word technology refers to. Two important things are overlooked here. One concerns general philosophical methodology, the other the particular situation of philosophy of technology. As for the former: if Heidegger's essay "The question concerning technology" is interpreted as a contribution to the philosophy of technology, this gets the order wrong. Heidegger wished to lay bare a certain phenomenon, and he felt justified to refer to that phenomenon as *Technik*. The phenomenon comes first, and is Heidegger's philosophical discovery, which makes him in a sense master of it. That others use the term in a different sense is, to Heidegger, an aspect of the phenomenon.

To be insensitive to this *modus operandi* in philosophy is bound to cause problems. The very same thing—although on a smaller scale—can be seen in how Ferré in his *Philosophy of Technology* discusses the work of Ellul. Ferré distinguishes four "problem areas" for the philosophy of technology, one of which is methodology. He then writes: "Some theorists hold that technology simply *is* methodology," and indicates in a footnote that Jacques Ellul is such a theorist. Here Ferré assumes that when Ellul equates "*La technique*" with methodology—Ellul defined it as "the totality of methods rationally arrived at and having absolute efficiency in every field of human activity"—Ellul's term "*technique*" refers to what the word "technology" as used by Ferré refers to. *Quod non*. Like Heidegger, Ellul claimed to have laid bare a fundamental phenomenon, a phenomenon which he felt justified to refer to as "*la technique*." No identification with common words is intended or should be inferred.

Once this is recognized, Heideggerian *Technik* and Ellulean *technique* cannot retain their benchmark status in the field of philosophy of technology. Neither Heidegger nor Ellul saw their work as contributing to philosophy of technology and both doubted the relevance of such a field, to put it mildly. Both in fact rejected the idea that their work belonged to philosophy at all, as long as that term refers to established academic philosophy. But even with Heidegger's *Technik* and Ellul's *technique* sidelined, the discipline still has a major problem concerning the meaning of technology, a problem that its practitioners seem increasingly happy to ignore rather than resolve.

From the start it was recognized that the term "technology" is used for different sorts of things. What these things are is fairly constant. Both Mitcham and Mackey (1972) and Kroes and Meijers (2000) distinguish three meanings: technology as a form of knowledge, as a set of operations or an activity, and as (a collection of) objects, in particular artifacts. Mitcham and Mackey suggested the use of three different terms for these three sorts of things—technology for the form of knowledge, technique for the activity, and technics for the objects—but without much confidence; they seemed to despair already at the start that their attempts to settle on a fixed meaning would be successful. But neither did they implement their own proposal; they continued to talk in terms of technology only, where the term stands for any and all of the things they have distinguished. Likewise, Kroes and Meijers first mentioned that the term technology can be found to have these three different meanings, only to ignore these differences for the remainder and to speak exclusively of technology.

Many authors either settle for something simpler or appear to take it for granted that a simple definition is the correct one. Each of the three candidates distinguished by Mitcham and Mackey and Kroes and Meijers can be found to serve this purpose. For Bunge (1985, 220), for example, technology simply is "the body of science-based technical knowledge," where the use of "technical" suggests a high degree of circularity. Verbeek (2005, 3 fn. 1) claims to follow "current usage" in taking "technology" to refer to "the specifically modern, "science-based" technological devices of the sort that begun to emerge in the last century" (by which I suppose the nineteenth century is meant). Again the use of "technological" suggests a high degree of circularity.

Of the simple type, activity definitions seem to be the most popular. They are preferred by engineers. Susskind, for example, equates technology to "man's efforts to satisfy his material wants by working on physical objects" (1973, 1). But philosophers also tend to prefer it. In his well-known book *Thinking through Technology*, Mitcham defined technology as "the making and using of artifacts" (1994, 1), and Joseph Pitt in *Thinking about Technology* curtly proposed the definition "technology is humanity at work" (1999, 11). No doubt these activity definitions were chosen because they seemed the most accurate: somehow they also involve the "knowledge" and "object" components. However, their being the most accurate goes together with their being the most problematic. Both "the making and using of artifacts" and "humanity at work" exclude very little. It is extremely difficult to imagine a situation where one is not "at work" or is not using some artifact. Technology, as Feenberg expresses it, is "the medium of daily life" (1999, vii).

At the other extreme of dealing with definitions, Langdon Winner has emphasized how the term technology is used "to talk about an unbelievably diverse collection of phenomena—tools, instruments, machines, organizations, methods, techniques, systems, and the totality of all these and similar things in our experience" and even claims that Ellul's "the totality of rational methods closely corresponds to the term technology as now used in everyday English [. . .] a vast, diverse, ubiquitous totality that stands at the center of modern culture [and which includes] a substantial portion of what we make and what we do" (1977, 8–9). This threatens to have the word mean everything and nothing. He therefore proceeded by proposing a few "technical" terms that each cover an aspect of the monstrous technology concept—apparatus, technique, organization, network. However, just like Mitcham and Mackey before him, he failed to implement his own proposal and talked in terms of technology for the rest of his book. Apparently Winner, Mitcham and Mackey, Kroes and Meijers, and undoubtedly many others as well, cannot see how to avoid using "technology" as an umbrella term that refers to a vague totality of activities, artifacts and knowledge, something that Ihde (1993a, 3) was careful to specify not more precisely than as a "phenomenon." As a result, as noted in the introduction, different claims about technology become incomparable.

It seems to me undeniable that this situation cannot be left like this if philosophy of technology is to prosper as a discipline. It also seems plausible to me that this situation encourages—certainly is incapable of discouraging—the eclecticism of importing widely diverging philosophical approaches and traditions into the field. However, it is itself a greater problem than this eclecticism. In order to achieve some progress, the concept of technology needs to be "tamed" first. In the next section I sketch a way to achieve this.

4. (Re)Structuring the Philosophy of Technology

How can philosophy of technology receive a clearer identity and overcome its state of being fragmented? What will not work is an attempt to unite the troops by waving a particular philosophical banner—either the banner of analytic philosophy or of continental philosophy. That distinction may well be approaching the end of its career. What philosophy needs is clarity of concepts and arguments, and although analytic philosophers may at times have suggested that only analytic philosophy is capable of delivering these, no philosophical approach can claim to own them. What is important is, first of all, a rough consensus on what unites the totality of interests and the work done into a single field—basically, how the concept of technology is understood and what are the basic problems and questions with respect to it. But if the philosophy of technology is to remain a single field which, at the same time, offers a place for all forms of questioning technology that philosophers have undertaken in the past half-century, a division into a small number of subfields seems desirable. In my view the following threefold division can serve as a starting point.

1. *First philosophy of Technology*. I choose this term to indicate the subfield where the basic concepts and basic statements of the field are investigated. Among the basic concepts is first of all that of an artifact, and more precisely a technical

artifact. Closely related are the concepts of function and the means-ends relation. Additionally the investigation of fundamental relations, foremost *making*—bringing into existence—and *using*, belongs here. Next to the grounding concepts and relations, the subfield of First Philosophy would investigate the character of statements and the elementary forms of reasoning central to technology. Ilkka Niiniluoto (1993) proposed statements of the form, "If one wants to achieve X, one should do Y"—statements forming a type that are called, after Von Wright (1963), 'technical norms'—as statements that are fundamental to applied research or design science. Remarkably, and unbeknownst to Niinililuoto and Von Wright, already a full century earlier Fred Bon (1898) described as 'philosophy of technology' the area of normative philosophy structured by statements of the form 'What should be done in order to achieve X?' A subfield of First Philosophy of Technology is where such fragmented attempts can come together and be systematically developed into a coherent framework to serve our thinking about practical action in the broadest sense.

2. *Philosophy of Engineering*. Contained in technology is the practice of making technical artifacts and artifactual systems. This includes everything from designing to manufacturing, implementing and even operating and maintaining. Within technology, only engineering can plausibly be seen as a practice of its own, similar to science. The two practices of science and engineering pervade one another to a high degree, but they remain distinct. The setting of goals, the processes of decision-making, the organization of work are all both more prominently social in character, and their social character is of greater significance to the practice than is the case in science. It is also more societal in character, that is, open to society as the broad environment into which the practice is embedded. Especially as regards the latter the difference to science is huge. One could say that engineering is much less master of the criteria and considerations central to it—effectiveness, efficiency, optimality—than science. One of the most astonishing aspects of the historical development of the philosophy of technology since the 1960s is that a (sub)field dedicated to Philosophy of Engineering has been extremely slow in developing. Perhaps the towering presence of the field of philosophy of science, which has hosted significant work relevant to the philosophy of engineering—for example Niiniluoto —may be one of the reasons for this, but the 'image' of philosophy among engineers, and the absence of any tradition of shared interests between philosophy and engineering, as it exists for philosophy and science, probably did not help either. However, I would argue that philosophy of technology has no future if it is not going to contain Philosophy of Engineering as a subfield.
3. *Philosophy of Technologies*. This is the subfield that studies the role of technology as implemented in society in the form of technologies: designed artifacts in use connected to a background of other artifacts. It addresses both how the use made of concrete technologies leaves its mark on society and culture and how technologies themselves are (re)shaped by the way they are used in society and by the effects this use helps cause.[11] The transformation force that technology

> exercises, increasingly so at the pace at which it develops, on the structure of society and the way that people live their lives—and on the lives that people can live—is such that philosophy of technology will be felt to be of no relevance whatsoever if it discards studying these aspects. And in fact it has never been seriously proposed that it should do so. Much criticism levelled at earlier work has argued it that if philosophy is to contribute in a meaningful and worthwhile way to the understanding of and resolution of the many issues and problems in this area, it should first—or minimally in parallel—develop an understanding of what technology is and how it comes into being prior to ending up as a social given. A discipline where Philosophy of Technologies coexists in a well-balanced way with First Philosophy of Technology and Philosophy of Engineering is a starting point for bringing this about.

Even if there is no substance to the idea that there are analytic and continental variants of philosophy of technology, still the distinction between them as variant ways of doing philosophy is a real one. Each comes with its specific types of expertise and its characteristic weaknesses and blind spots. We may therefore expect the distribution of philosophers who roughly identify with either of these variants, if only as their educational background, to be far from uniform. Undoubtedly, philosophers raised in the analytic tradition will feel perfectly at home in the subfield of First Philosophy. It has seen some major activity in the past two decades by philosophers from the Netherlands, for example in the form of "The dual nature of technical artifacts" research program (Kroes & Meijers 2006) and a follow-up project dedicated to the metaphysics of artifact kinds (Franssen et al. 2014). This work can be placed in the tradition of analytic philosophy, and indeed clarity and precision—although not necessarily through formal logic—seem of crucial importance here. But an emphasis on clarity and precision in no way closes off an area for certain topics or approaches. An elaboration of the mediation view of artifacts adopted by Ihde and especially Verbeek, which until now has remained rather sketchy, I would consider a key contribution to the subfield of First Philosophy of Technology.

With respect to the subfield of Philosophy of Engineering, the major challenge is to develop this into a recognizable and coherent enterprise. Only since a decade or so has work been done to articulate this subfield.[12] Writings that contain an open invitation to do so, with a rough sketch of what it would deal in, have been lying in wait for up to several decades (e.g. Simon 1969, Vincenti 1990). Since science and engineering are both practices, and very interwoven at that, the philosophy of science will function as a benchmark of sorts. Taking a comparative approach would serve to speed up the development of the philosophy of engineering. As this subfield is particularly close to philosophy of science, given how science-based modern engineering is and given how close engineering education is to science education, those who consider themselves, broadly, analytic philosophers will feel at home in this subfield as well. Important work in this comparative vein has already been done: in addition to the works mentioned in n. 12, for example, Houkes (2009) on engineering knowledge and Zwart and De Vries

(2016) on the nature of engineering research. The subfield clearly has many points of contact with the subfield of First Philosophy through concepts as artifact and function, which are central to many aspects of the practice. However, the subfield can also be expected to have clear points of contact with the subfield of Philosophy of Technologies. Engineering is an activity that overwhelmingly takes place in business firms and this fact must be taken into consideration. How individual and societal wants and needs reach engineers and firms, either directly or mediated by all sorts of public agents, how the environment, structured by democracy and competition among firms in the market, greatly affects the way the practice of engineering is organized, structures how innovation happens and how it does not happen, and has a say in which products are designed and which technologies are developed. These are all matters that are of importance for the Philosophy of Engineering, for example concerning how it models engineering decision making. Philosophy of Engineering, therefore, should not be considered of exclusive interest to analytic philosophers. Bucciarelli (1994) has adopted an approach that leans somewhat to constructivism in being "ethnological" but which offers precisely the sort of empirical work that philosophical analyses must take into account. His emphasis on the business firm as the default environment for engineering, and his analysis of this environment in terms of the concept of "object world," defined through engineering disciplines, make his work a meeting place for analytic and narrative approaches.

The third subfield, that of Philosophy of Technologies, is where most of the work done until now must be placed. Insofar as this work can qualify as philosophical, and aims to be so qualified, it represents a wide spectrum of different perspectives. In that spectrum, however, the perspective of analytic philosophy is not particularly prominent. Nevertheless, the emphasis on conceptual clarity and argumentative precision that analytic philosophy strives for are definitely in need here, as this subfield faces a number of challenges. How these challenges are dealt with, and whether they are acknowledged in the first place, will play a significant role in the coming to adulthood of philosophy of technology—as judged from the vantage points of philosophy, of engineering, and of the human and social sciences.

I have room here to briefly discuss only one challenge. Philosophers are generally prone to an individualistic bias and tend to ignore the mechanisms of aggregation that "generate" society and social phenomena from the actions of individuals, and the gap that separates micro-level phenomena—the level of individual behavior—from macro-level phenomena—the level of social structure.[13] Philosophical concepts are defined overwhelmingly with reference to the deliberating and acting individual. Most of these concepts cannot be transferred to the aggregate level. To talk in terms of "humanity," "mankind," "man," is to put up smoke screens—they hide from view that at the aggregate level there are no subjects or agents but things that happen. For example, in philosophy of technology one regularly encounters the statement that technology is instrumental—if only as the standard view, which is then criticized. However, if it is criticized, it is not for the right reason: critique is directed at the view that technology is *merely* instrumental, not the idea that one can think of technology as instrumental in the first place. But precisely the latter is problematic. There is a great risk here of falling victim to category mistakes—a notion from analytic philosophy introduced by Ryle (1949)—which is eminently relevant to Philosophy of Technologies. Individual artifacts are correctly looked upon as instrumental: they are produced to be used in well-defined environments to serve specific and concrete purposes, and that is generally how they are in fact used. However, this does not make implemented technologies instrumental in this sense. Given the enormously distributed way in which technology is developed and implemented, the total configuration of systems that consist of both artifacts and humans engaged with them is the result of historical accident and changes faster than anyone can record and cannot be said, in its totality, to serve any particular purpose. No such purpose has ever been defined or conceived, nor is a subject available who can be said to entertain it. To state that technology "is for" increasing human welfare is whistling in the dark. That technology offers a reservoir of means, "technologies," from which people select what suits them to serve their private ends and governments what suits them to serve public ends, articulated in whatever way, does not mean that technology can be seen as a global instrument which "humanity" uses to a purpose. "Humanity" is the mere receptor of the net result of the existence of technology. What lives people live, can live, and would want to live is determined at any moment by the total state of the world, including the state of technology.

This issue is of particular relevance to the ethics of technology. Over the past decades, the assessment of the way technology influences society and culture and individual lives has increasingly been made subject to scrutiny from the perspective of ethics. Ethics of technology is now an accepted term, especially within engineering education. Ethics, however, as a philosophical field of study, takes the acting individual as starting point. Ethics judges, prescribes, and assesses the actions, choices, and attitudes of individual human beings. It is highly contentious, to say the least, to what extent any of it can remain valid once we start to ascend levels of aggregation. The "problem of many hands" is notoriously ubiquitous in engineering and technology (see e.g. Van de Poel, Royakkers, and Zwart 2015). However, we need only to look at the work of Margaret Gilbert (1989) on plural subjects—exemplary analytic philosophy—to see the amount of detail that goes into elevating intentions to the level of small groups such that ethical concepts like responsibility can be given a meaning at the aggregate level. The philosopher's choice of ethics as the conceptual framework for approaching issues concerning the assessment and evaluation of technology's role in society runs a serious risk of obscuring rather than clarifying. The general domain of philosophical reflection on values and normativity is in fact separated—inevitably in view of the complexities caused by aggregation—into ethics for the individual level and political philosophy for the societal level.

To be sure, this separation cannot be total, since individuals are also members of society—something that people were already keenly aware of in Antiquity, as we can find the conflict between individual morality and societal obligation already treated in Greek tragedy, e.g. *Antigone* of Euripides. There are authors whose work straddles this bifurcation, in particular John Rawls, but the point of reference is still formed by individual choices, and the problem of how to assess at higher levels of aggregation, and whether this is possible at all using individual-level concepts, is not addressed. Ethics of technology should beware of trying to hew a path through a jungle that has already been charted over the course of several centuries by the discipline of political theory. There are philosophers of technology who approach Philosophy of Technologies primarily from a political perspective rather than an ethical one—notably Winner and Feenberg—but still the ethical approach currently seems to dominate.

5. Conclusion

Given the vastness of the phenomenon of technology, clarity, precision, and analysis will be indispensable not only if the philosophy of technology is to be acknowledged as having something relevant to contribute in the totality of "concerns" that people have with respect to technology, but also if it is to be acknowledged as making its contribution as philosophy. Clarity, precision, and analysis are virtues that analytic philosophy in particular claims for itself. This is not idle talk. See for instance the notion of a category mistake, or the array of concepts, like "practice," that are now available to bring structure to an extremely wide-ranging concept as "technology," all products of analytical philosophy's focus on conceptual analysis. However, analytic philosophy is itself quite vulnerable to an overemphasis on the intentionally acting individual and its conceptual outlook. This is one of the main reasons why it received criticism both from continental philosophy and constructivism. There is a considerable amount of truth in the complaint from the social-scientific approach (of which social constructivism is the most radical representative) that philosophy, in particular philosophy of science, is exclusively interested in rational reconstruction and is therefore inclined, or even condemned, to write Whig history in the service of science. The problem is, however, that the social science which constructivism introduced as an alternative is shallow and impoverished, and occasionally seems even to have been adopted primarily for its potential to upset philosophers rather than its potential to clarify and explain social processes. Still, philosophical understanding must be distinguished from (social-) scientific understanding. Philosophy is the unique discipline in which normative questions take center stage: questions concerned with values and meaning—conceptual meaning as well as life-guiding meaning. But philosophy itself has no methodology for penetrating social phenomena. It can only contribute by penetrating the methods and theories that the humanities and social sciences use to penetrate social phenomena. Just as in the subfield of Philosophy of Engineering philosophers will have to cooperate closely with engineers and heed both what they say they do and what they actually do, in the subfield of Philosophy of Technologies philosophers will have to cooperate closely with historians and social scientists and to calibrate their interpretations against the findings of these disciplines.

Notes

1. See Critchley (2001, 38). Critchley mentions an early use by John Stuart Mill but he used it in the literal, and as such quite neutral, sense to refer to the work of philosophers on the European continent in general. It is unclear how wide this usage was and whether it informed modern usage.
2. Prior to Ihde, Mitcham and Mackey (1972) had included Heidegger as one of the many authors who had addressed "the philosophical problems of technology." Possibly because Heidegger's "The Question Concerning Technology" was not yet available in English, the volume "attended to" Heidegger in the form of an essay by Hood in which the "Aristotelian view" of technology and the "Heideggerian view" were contrasted and the latter was recommended for the understanding of modern technology.
3. Glock (2008) offers a well-researched, book-length overview; Raatikainen (2013) is also helpful. As far as continental philosophy is concerned, apart from Critchley (1998, 2001) I am acquainted only with Mulligan (1991) as sketches of a history. Critchley offers broad and narrow conceptions of continental philosophy as possibilities but prefers a broad conception, whereas I prefer, as a working hypothesis, a narrow conception. A different but similarly narrow conception has been entertained by Dummett (1993).
4. I see the contrast between Husserl's view (in *The Crisis of European Sciences*, 1970) that his work completed Western philosophy and Heidegger's resolve to step out of the circle of "completed metaphysics" and even to destroy metaphysics as crucial. My conception of core continental philosophy is a narrow one, therefore, but a justification of this conception has to be undertaken elsewhere.
5. In "The Basic Problems of Phenomenology," a course read in the summer semester of 1927, roughly a summary of *Being and Time*, the distinction between *Vorhanden* and *Zuhanden* is lacking (Heidegger 1982, 161–170).
6. The distinction between ready-to-hand and present-at-hand is not a dichotomy. Heidegger suggests that nature and natural entities outside of human reach "are" in yet another way.
7. Dominic Smith, whose *Exceptional Technologies: A Continental Philosophy of Technology* (2018) is the only explicit attempt at continental philosophy of technology that I know of, has a similar lackadaisical approach to how contemporary philosophy in the continental tradition relates to its foundational themes: to him what is common to all continental philosophy is "a sense of the transcendental."
8. Ironically, social constructivism's pedigree can be traced straightforwardly to philosophy, and to the heart of analytic philosophy at that, since it was Peter Winch's interpretation of Wittgenstein's *Philosophical investigations* that was a major input for David Bloor's formulation of the strong program.
9. The 'nature' underlying it is of course that formed by humans in association, so the better term would be 'socio-culturalism.'
10. In (Franssen 2009) I began my discussion of analytic philosophy of technology by saying there is no such thing.
11. Both Verbeek (2005, 2011) and Smith (2018) address and question primarily technologies, not technology—which has motivated me to choose the term 'philosophy of technologies.'
12. A major step has been taken with the biennial conferences of the Forum for Philosophy and Engineering, held since 2010, following up two international Workshops on Philosophy and Engineering in 2007 and 2008 and resulting in several edited volumes with conference

papers (Michelfelder, McCarthy, and Goldberg 2013; Michelfelder, Newberry, and Zhu 2017; Fritzsche and Oks 2018). Apart from this see also (Bulleit et al. 2015).

13. An excellent introduction to these problems, addressed already in the book's title, is Schelling's *Micromotives and Macrobehavior* (1978).

References

Amsterdamska, Olga. 1990. "Surely You Are Joking, Monsieur Latour!" *Science, Technology & Human Values* 15: 495–504.

Bloor, David. 1999. "Anti-Latour." *Studies in the History and Philosophy of Science* 30: 81–112.

Bon, Fred. 1898. *Über das Sollen und das Gute: Eine begriffsanalytische Untersuchung*. Leipzig: Engelmann.

Bucciarelli, Louis L. 1994. *Designing Engineers*. Cambridge, MA/London: MIT Press.

Bulleit, William, Jon Schmidt, Irfan Alvi, Erik Nelson, and Tonatiuh Rodriguez-Nikl. 2015. "Philosophy of Engineering: What It Is and Why It Matters." *Journal of Professional Issues in Engineering Education and Practice* 141(3): 1–25.

Bunge, Mario. 1985. *Treatise on Basic Philosophy*. Volume 7. Philosophy of Science and Technology, Part II. Dordrecht/Boston, MA/Lancaster: D. Reidel.

Critchley, Simon. 1998. "Introduction: What Is Continental Philosophy?" In *A Companion to Continental Philosophy*, edited by Simon Critchley and William R. Schroeder, 1–17. Malden, MA/Oxford: Wiley-Blackwell.

Critchley, Simon. 2001. *Continental Philosophy: A Very Short Introduction*. Oxford: Oxford University Press.

Dessauer, Friedrich. 1927. *Philosophie der Technik: Das Problem der Realisierung*. Bonn: Friedrich Cohen.

Dummett, Michael. 1993. *Origins of Analytic Philosophy*. London: Duckworth.

Ellul, Jacques. 1954. *La Technique: L'Enjeu du siècle*. Paris: Armand Colin. English edition: *The Technological Society*. 1964. Translated, with an introduction, by John Wilkinson. New York: Alfred A. Knopf.

Engelmeier, Petr K. 1912. *Filosofija techniki*. 4 vol. Moskva: A. A. Levenson.

Feenberg, Andrew. 1999. *Questioning Technology*. London: Routledge.

Ferré, Frederick. 1988. *Philosophy of Technology*. Englewood Cliffs, NJ: Prentice Hall.

Franssen, Maarten. 2009. "Analytic Philosophy of Technology." In *A Companion to the Philosophy of Technology*, edited by Jan Kyrre Berg Olsen Friis, Stig Andur Pedersen, and Vincent F. Hendricks, 184–188. Malden, MA/Oxford: Wiley-Blackwell.

Franssen, Maarten, and Stephan Koller. 2016. "Philosophy of Technology as a Serious Branch of Philosophy: The Empirical Turn as a Starting Point." In *Philosophy of Technology after the Empirical Turn*, edited by Maarten Franssen, Pieter E. Vermaas, Peter Kroes, and Anthonie W. M. Meijers, 31–61. Dordrecht: Springer.

Franssen, Maarten, Peter Kroes, Thomas A.C. Reydon, and Pieter E. Vermaas, eds. 2014. *Artefact Kinds: Ontology and the Human-made World*. Dordrecht: Springer.

Fritzsche, Albrecht, and Sascha Julian Oks, eds. 2018. *The Future of Engineering: Philosophical Foundations, Ethical Problems, and Application Cases*. Cham: Springer.

Gehlen, Arnold. 1957. *Die Seele im technischen Zeitalter: Sozialpsychologische Probleme in den industriellen Gesellschaft*. Reinbek: Rowohlt.

Gilbert, Margaret. 1989. *On Social Facts*. London: Routledge.

Glock, Hans-Johann. 2008. *What Is Analytic Philosophy?* Cambridge: Cambridge University Press.

Heidegger, Martin. 1962. *Being and Time*. Translated by John Macquarrie and Edward Robinson. Oxford/Cambridge, MA: Blackwell. Translation of *Sein und Zeit*. 1927. Halle an der Saale: Max Niemeyer.

Heidegger, Martin. 1973. "Overcoming Metaphysics." In *The End of Philosophy*, translated by Joan Stambaugh, 84–110. New York: Harper & Row. Translation of "Überwindung der Metaphysik." 1954. In *Vorträge und Aufsätze*, 71–99. Pfüllingen: Günther Neske.

Heidegger, Martin. 1977. "The Question Concerning Technology." In *The Question Concerning Technology and Other Essays*, translated and introduced by W. Lovitt, 3–35. New York: Harper & Row. Translation of "Die Frage nach der Technik." 1954. In *Vorträge und Aufsätze*, 13–44. Pfüllingen: Günther Neske.

Heidegger, Martin. 1982. *The Basic Problems of Phenomenology*. Translated by Albert Hofstadter. Bloomington, IN: Indiana University Press. Translation of "Die Grundprobleme der Phänomenologie." 1975. In *Gesamtausgabe*, vol. 24, edited by Friedrich-Wilhelm von Herrmann. Frankfurt am Main: Vittorio Klostermann.

Houkes, Wybo. 2009. "The Nature of Technological Knowledge." In *Philosophy of Technology and Engineering Sciences*, edited by Anthonie Meijers, 309–350. Amsterdam: North-Holland.

Husserl, Edmund. 1970. *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. Translated by David Carr. Evanston: Northwestern University Press. Translation of *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*. Edited by Walter Biemel. 1954. Den Haag: Martinus Nijhoff.

Ihde, Don. 1979. *Technics and Praxis*. Dordrecht/Boston, MA/London: D. Reidel.

Ihde, Don. 1993a. *Philosophy of Technology: An Introduction*. New York: Paragon House.

Ihde, Don. 1993b. *Postphenomenology: Essays in the Postmodern Context*. Evanston, IL: Northwestern University Press.

Jünger, Friedrich Georg. 1949. *Die Perfektion der Technik*. Frankfurt am Main: Vittorio Klostermann.

Kapp, Ernst. 1877. *Grundlinien einer Philosophie der Technik: Zur Entstehungsgeschichte der Cultur aus neuen Gesichtspunkten*, Braunschweig: George Westermann; English edition, 2018. *Elements of a Philosophy of Technology: On the Evolutionary History of Culture*. Translated by Lauren K. Wolfe. Minneapolis, MN: University of Minnesota Press.

Kranzberg, Melvin. 1966. "Toward a Philosophy of Technology." *Technology and Culture* 7: 301–302.

Kroes, Peter, and Anthonie Meijers. 2000. "Introduction: A Discipline in Search of Its Identity." In *The Empirical Turn in the Philosophy of Technology*, edited by Peter Kroes and Anthonie Meijers, xvii–xxxv. London: JAI Press.

Kroes, Peter, and Anthonie Meijers, eds. 2006. *Studies in History and Philosophy of Science* 37: 1–158. Special issue dedicated to the dual nature of technical artefacts.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Milton Keynes: Open University Press.

Latour, Bruno. 1988. *The Pasteurization of France*. Translated by Alan Sheridan and John Law. Cambridge, MA/London: Harvard University Press. Translation of *Les Microbes: guerre et paix*, suivi de *Irréductions*. 1984. Paris: A.M. Métailié.

Michelfelder, Diane P., Natasha McCarthy, and David E. Goldberg, eds. 2013. *Philosophy and Engineering: Reflections on Practice, Principles and Process*. Dordrecht: Springer.

Michelfelder, Diane P., Byron Newberry, and Qin Zhu, eds. 2017. *Philosophy and Engineering: Exploring Boundaries, Expanding Connections*. Cham: Springer.

Mitcham, Carl. 1994. *Thinking through Technology: The Path between Engineering and Philosophy*. Chicago: University of Chicago Press.

Mitcham, Carl, and Robert Mackey. 1972. "Introduction: Technology as a Philosophical Problem." In *Readings in the Philosophical Problems of Technology*, edited by Carl Mitcham and Robert Mackey, 1–30. New York: The Free Press.

Mulligan, Kevin. 1991. "Introduction: On the History of Continental Philosophy." *Topoi* 10: 115–120.

Niiniluoto, Ilkka. 1993. "The Aim and Structure of Applied Research." *Erkenntnis* 38: 1–21.

Peterson, Martin. 2012. "Three Objections to Verbeek." *Philosophy and Technology* 25: 619–626.

Pitt, Joseph C. 1999. *Thinking about Technology: Foundations of the Philosophy of Technology*. New York/London: Seven Bridges Press.

Raatikainen, Panu. 2013. "What Was Analytic Philosophy?" *Journal for the History of Analytical Philosophy* 2 (2): 11–27.

Rapp, Friedrich. 1981. *Analytical Philosophy of Technology*. Translated by Stanley R. Carpenter and Theodor Langenbruch. Dordrecht/Boston, MA/London: D. Reidel. Translation of *Analytische Technikphilosophie*. 1978. Freiburg/München: Alber.

Ryle, Gilbert. 1949. *The Concept of Mind*. London: Hutchinson.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior*. New York: W.W. Norton.

Schröter, Manfred. 1934. *Philosophie der Technik*. Berlin/München: Oldenbourg.

Simon, Herbert A. 1969. *The Sciences of the Artificial*. Cambridge, MA/London: MIT Press.

Simondon, Gilbert. 1958. *Du mode d'existence des objets techniques*, Paris: Aubier. English edition, *On the Mode of Existence of Technical Objects*. Translated by Cécile Malaspina and John Rogove. 2017. Minneapolis, MN: Univocal Publishing.

Smith, Dominic. 2018. *Exceptional Technologies: A Continental Philosophy of Technology*. London: Bloomsbury Academic.

Susskind, Charles. 1973. *Understanding Technology*. Baltimore, MD: Johns Hopkins University Press.

Van de Poel, Ibo, Lambèr Royakkers, and Sjoerd D. Zwart. 2015. *Moral Responsibility and the Problem of Many Hands*. London: Routledge.

Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. Translated by Robert P. Crease. University Park, PA: Pennsylvania State University Press. Translation of *De daadkracht der dingen: Over techniek, filosofie en vormgeving*. 2000. Amsterdam: Boom.

Verbeek, Peter-Paul. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*, Chicago/London: University of Chicago Press.

Verbeek, Peter-Paul. 2012. "The Irony of Humanism: On the Complexities of Discussing the Moral Significance of Things." *Philosophy and Technology* 25: 626–631.

Vincenti, Walter G. 1990. *What Engineers Know and How They Know It: Analytical Studies From Aeronautical History*. Baltimore, MD: Johns Hopkins University Press.

Von Wright, Georg Henrik. 1963. *Norm and Action: A Logical Enquiry*. London: Routledge & Kegan Paul.

Winner, Langdon. 1977. *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. Cambridge, MA/London: MIT Press.

Winner, Langdon. 1993. "Upon Opening the Black Box of Technology and Finding It Empty: Social Constructivism and the Philosophy of Technology." *Science, Technology & Human Values* 18: 362–378.

Zschimmer, Eberhard. 1914. *Philosophie der Technik: Vom Sinn der Technik und Kritik des Unsinns* über *die Technik*. Jena: Eugen Diederichs.

Zwart, Sjoerd D., and Marc J. de Vries. 2016. " Methodological Classification of Innovative Engineering Projects." In *Philosophy of Technology after the Empirical Turn*, edited by Maarten Franssen, Pieter E. Vermaas, Peter Kroes, and Anthonie W. M. Meijers, 219–248. Dordrecht: Springer.

CHAPTER 5

WHENCE AND W(H)ITHER TECHNOLOGY ETHICS

DON HOWARD

1. Introduction: Technology Ethics for the Twenty-first Century

Technology ethics, by contrast with its more wide ranging parent discipline, the philosophy of technology, is a still developing field of study that is not yet fully established as an independent, academic sub-discipline, as judged by such sociological markers as a dedicated professional organization and the recruitment of college and university faculty with technology ethics as the stipulated area of specialization. Of late, interest in the field, both inside and outside of the academy, is picking up. But growing the community of scholarship on technology ethics, making that scholarship still more sophisticated, and bringing the scholarship into conversation with engineers, entrepreneurs, corporate executives, regulators, legislators, and consumers are compelling needs in a world in which the ever-accelerating pace of technological innovation poses ever more problems, and ever more serious problems, regarding the ethical impacts of new technologies. Those of us who work in this space must be more intentional in shaping the future of this still emerging and evolving field. In doing so we have to ask ourselves, what are our primary goals for technology ethics as a discipline and by what criteria we will judge our success in achieving those goals? I would guess that most of us want technology ethics to be not a detached academic specialty but an engaged body of scholarship that does its part to make ours a better world.

If our goal is engaged scholarship aiming to promote human flourishing, then we must ask ourselves how the scholarship that we produce can best serve that end. The more specific question that I want to pose is whether we do this better with scholarship focused mainly on the identification and mitigation of risk (what I might term "monitory" scholarship), or also with scholarship that seeks out and promotes responsible technological innovation that promises moral gain (what I might term "amelioratory" scholarship)? My concern is that, for reasons having to do with the circumstances of its birth and early development, technology ethics as a field of study has foregrounded monitory scholarship, and that its having done so has compromised its prospects for constructive dialogue with the engineers, executives, regulators, legislators, and consumers who should be an important part of its primary audience. Or, to say it more plainly, if all that we do is wag a finger, say "tsk, tsk," and tell the engineers what they are not allowed to do, then we risk making our scholarship not irrelevant, but unread and, so, inconsequential in the very quarters where it is most needed.

2. A Maculate Conception? The Birth of Technology Ethics

Technology ethics stands in a complex relationship with several nearby fields of scholarship, including the philosophy of technology, the history of technology, science and technology studies (STS), the sociology of scientific knowledge (SSK), and the history and philosophy of science (HPS), along with newer groupings, such as the Consortium for Socially Responsible Philosophy of/in Science and Engineering (SRPoiSE) and more specialized fields like environmental ethics and medical ethics. It speaks many languages beyond the major European ones, but in the form in which it exists in the core, contemporary literatures, its origins are, for better or worse, distinctly European and North American.

In the twentieth century, serious philosophical reflection on technology begins with Friedrich Dessauer's 1927 book, *Philosophie der Technik* [*Philosophy of Technology*] (Dessauer 1927), and Lewis Mumford's 1934 *Technics and Civilization* (Mumford 1934; see also Mumford 1967–1970). If there is, however, one "Urquell," one original source from which the field that was to become technology ethics first emerged, it was Martin Heidegger's "The Question Concerning Technology" ["Die Frage nach der Technik"] (Heidegger 1953). A fixture even today on every undergraduate Science and Technology Studies (STS) or Science, Technology, and Values (STV) course syllabus, no one work has done more to set the tone and the valence of philosophical thinking about technology's impact on the world and human experience in a technologized world. Heidegger warns of specific risks, such as global annihilation via nuclear weapons, and laments the tendency of technology or our uncritical embrace of technology to transform the whole world, including humankind, into a "standing reserve," essentially the raw material with which technology works. But his deeper message, a nostalgic, neo-Romantic message, is that a technologically mediated relation of humankind to the world, what he calls "Enframing," precludes or impedes a more authentic and organic approach to the world through an openness to Being:

> The threat to man does not come in the first instance from the potentially lethal machines and apparatus of technology. The actual threat has already affected man in

> his essence. The rule of Enframing threatens man with the possibility that it could be denied to him to enter into a more original revealing and hence to experience the call of a more primal truth.
>
> (Heidegger 1953, 28)

The real risk is, thus, a metaphysical one. Yes, there is more than a little irony in the fact that this critique of technology's threat to the very essence of humankind was penned by a philosophical apologist for the ideology of Nazism and anti-Semitism that pioneered technologized murder on a scale theretofore unknown in human history. But bracket that. The more important point is that this founding text is born out of a profoundly anti-modern, anti-Enlightenment intellectual project, one deeply skeptical of any of the otherwise common philosophical efforts to link science and technology to freedom, democracy, and material progress, be they the naïve scientism of liberal, democratic political theory, the "scientific world view" championed by Vienna Circle logical empiricists, or Marxist "scientific materialism."

Another thinker of the same era who, while less well known today, had at that time a comparable impact on the philosophy of technology in its early years and nascent technology ethics was the French philosopher, sociologist, and theologian, Jacques Ellul, whose 1954 book, *La Technique: L'Enjeu du siècle* (Ellul 1954) found a wide readership when it appeared in an English translation ten years later as *The Technological Society* (Ellul 1964). Ellul's intellectual heritage was markedly different from Heidegger's, the chief influences on his thinking being Karl Marx, Søren Kierkegaard, and Karl Barth. His politics, likewise, differed from Heidegger's. He was active in the French resistance during World War II and was named one of the "Righteous Among the Nations" at Yad Vashem in 2001. Nonetheless, his technology critique converged with Heidegger's in several respects. Ellul viewed technological dominance as just one especially threatening instance of the broader phenomenon, characteristic of modernity, of the penetration of what he termed "technique" in almost every domain of human experience, "technique" being a sociocultural formation that valorizes rationality and efficiency. Ellul's concept of "technique" bears comparison with Heidegger's notion of "Enframing." Also like Heidegger, Ellul argued that technology had fundamentally altered humanity's relation with nature, modern science having begun the desacralization of nature as far back the seventeenth century:

> The world that is being created by the accumulation of technical means is an artificial world and hence radically different from the natural world.
>
> It destroys, eliminates, or subordinates the natural world, and does not allow this world to restore itself or even to enter into a symbiotic relation with it. The two worlds obey different imperatives, different directives, and different laws which have nothing in common. Just as hydroelectric installations take waterfalls and lead them into conduits, so the technical milieu absorbs the natural. We are rapidly approaching the time when there will be no longer any natural environment at all.

> When we succeed in producing artificial *aurorae boreales*, night will disappear and perpetual day will reign over the planet.
>
> (Ellul 1964, 79)

In effect, science and technology have become the new sacred in the modern world.

Though Ellul and Heidegger share a neo-Romantic nostalgia for a lost, pre-industrial world in which humankind stood in a more authentic relationship with nature, Ellul is not at all as resigned and reactionary as Heidegger. He writes as a sociologist seeking mainly to understand our modern, technologized, human condition. He documents what he takes to be the likely further development and domination of the technical imperative, but he does this not in a spirit of total despair and powerlessness:

> In the modern world, the most dangerous form of determinism is the technological phenomenon. It is not a question of getting rid of it, but, by an act of freedom, of transcending it. How is this to be done? I do not yet know. That is why this book is an appeal to the individual's sense of responsibility. The first step in the quest, the first act of freedom, is to become aware of the necessity . . . . [B]y grasping the real nature of the technological phenomenon, and the extent to which it is robbing him of freedom, [man] confronts the blind mechanisms as a conscious being.
>
> (Ellul 1964, xxxiii)

There is more than a hint of Marx, Kierkegaard, and Barth in the way in which Ellul here puts human freedom and technological determinism into a dialectical relationship with one another. But while Ellul is not wholly resigned to a tragic fate of the total technological domination of nature and human life, he shares with Heidegger an essentially tragic reading of the role of technology in human affairs.

A third voice that shaped emergent philosophy of technology and technology ethics in the mid-twentieth century was that of neo-Marxist, Frankfurt School critical theory. Herbert Marcuse and Max Horkheimer were the movement's most prominent refugee representatives in the United States in the 1930s, where they re-established the Institut für Sozialforschung as the Institute for Social Research at Columbia University. A critique of the epistemology of modern scientific reason had played a central role in the work of the Frankfurt School for some time, epitomized by Horkheimer's seminal, 1937 essay, "Traditionelle und kritische Theorie" ["Traditional and Critical Theory"] (Horkheimer 1937), where Horkheimer contrasted conventional scientific theory, which sought only understanding, explanation, and through them, control of the material world, with transformative, critical, social theory. Thinking out the implications for the kind of reason embodied specifically in technology, Marcuse introduced the notion of "technological rationality" in his 1941 essay on "Some Social Implications of Modern Technology" (Marcuse 1941), and that was followed by Horkheimer's articulation of the kindred notion of "instrumental reason" in his 1947 book, *The Eclipse of Reason* (Horkheimer 1947).

On Marcuse's analysis, technological rationality is the ironic outgrowth of the "individualistic rationality" that theorized and legitimated the middle-class revolutions of the eighteenth and nineteenth century, the revolutions that ushered in the industrial age that now demands its own form of reason, one repudiating the very notions of freedom and individualism in the name of which those revolutions were fought. Technological rationality prioritizes efficiency and compliance:

> The idea of compliant efficiency perfectly illustrates the structure of technological rationality. Rationality is being transformed from a critical force into one of adjustment and compliance. Autonomy of reason loses its meaning in the same measure as the thoughts, feelings and actions of men are shaped by the technical requirements of the apparatus which they have themselves created. Reason has found its resting place in the system of standardized control, production and consumption. There it reigns through the laws and mechanisms which insure the efficiency, expediency and coherence of this system.
>
> (Marcuse 1941, 422)

This technological rationality is not just a thing of the mind. It is bound up with the material forces of production and reconfigures all social relationships to suit the needs of technology, including the creation of mass bureaucracy for the administration of both humans and machines.

As with Ellul, who was also a student of Marx, Marcuse and Horkheimer do not simply despair, however grim and realistic their vision of the totalizing tendencies of modern, technological societies, especially in the more horrific form they took in places like Nazi Germany. Following the lead of Marx, they recognize that, just as capitalism contains within itself the seeds of its own destruction, so, too, emancipatory opportunities might emerge from within a world shaped by technological rationality, if only critical reason can also be brought to bear on the problem and if public forms of organization focused on the genuine needs of humankind can be developed, which means massive, deliberate, progressive, governmental reform. Still, they recognized that the challenge was a daunting one and that the forces of progressive social and political reform were, then, not yet adequate to the task.

It was not as if all thinkers in the post-war era were equally dour in their assessments of the impact of technology on the natural world and human well-being. A thoughtful alternative view was put forward in C. P. Snow's 1959 Rede Lecture, "The Two Cultures and the Scientific Revolution" (Snow 1959). We wrongly remember the lecture as mainly a meditation on the challenges of cross-disciplinary communication between humanists and scientists. It was that, but Snow's main message was importantly different. For one thing, he was more concerned with relations between humanists and engineers, not scientists, and the reason for his concern was not a purely theoretical worry about the prospects for interdisciplinarity or cross-disciplinary communication. No, his real concern was that the world faced many, serious problems that would be far harder to solve were the humanists and the engineers to persist in talking at cross purposes. Why?

Snow characterized the engineer as the "cultural optimist," someone who believes that challenges can be met and problems solved by the application of science and reason. The "literary humanist," by contrast, he described as the "cultural pessimist," someone so overwhelmed by the tragic nature of the human condition as to be rendered powerless. One easily imagines that Snow had thinkers like Heidegger, Ellul, Horkheimer, and Marcuse in mind when thinking about the mindset of the humanist. Snow valued the insights of the humanists for their tempering the often naïve optimism of the engineer with a keen sense of the complexity of the world and the baleful impact of the presence in it of fallen, sinful humankind. But Snow was first trained as a chemist, and he cast his lot with the engineers, as was made clear in the lengthy essay, "A Second Look," that he appended to the original lecture in the 1963 expanded edition, where, among other things, he presents a passionate argument for why the nations of the West must invest heavily in training a vastly expanded, new generation of young engineers if we are to compete for global dominance with the Soviet Union. His point was not that we need to build more and better bombs, but that we need to build schools, hospitals, highways, power plants, and agricultural infrastructure in the developing world. Humanists must play a vital role in this enterprise, if only to help us understand the sometimes very different cultures to which we seek to extend a helping hand, but humanists cannot do that if they think that we are doomed, come what may, and that technology is the cause of our damnation. Snow was a Cold Warrior, but if we write off his argument as merely a propaganda exercise, then we miss the central point about which he mainly cared, which is that hand-wringing and another sip of absinthe will not fix a world at risk.

Snow's call to train a new generation of scientists and engineers who were equally well educated in history, literature, philosophy, anthropology, and the arts garnered a large and enthusiastic following among his technical colleagues around the world and among a younger generation of aspiring young scientists and engineers who wanted to put their brains to work making not weapons but a world of peace and prosperity. But too many of the humanists whose help Snow earnestly sought either ignored the argument or took it as another excuse to condescend to scientists and engineers who could not, from memory, quote long bits of Shakespeare, as when the eminent Cambridge literary scholar F. R. Leavis, condemned Snow as "portentously ignorant" of both literature and history (Leavis 1963, 28), calling him a "'public relations' man for Science" (Leavis 1963, 14).

It did not help that John Kennedy's Secretary of Defense, former Ford Motor Company president Robert McNamara, was eager to recruit a lot of smart, young technocrats, the "best and the brightest" (Halberstam 1972), to plan and execute with Kennedy's successor, Lyndon Johnson, a massive expansion of the US role in the Vietnam War. Kennedy, himself, had been more invested in recruiting the same kind of technical talent for the Peace Corps, the mission of which—international development and peace building—aligned more closely with Snow's vision of what culturally sophisticated and sensitive engineers could achieve. But that was not the mood in 1963 among the humanists to whom Snow reached a hand in peace.

No, the dominant mood on the humanist side of the academy in 1963 and among many in the broader public was otherwise. Marcuse's critique of technological rationality found a large and receptive audience when it was repackaged as a central theme in his widely-read book, *One-Dimensional Man* (Marcuse 1964), which appeared in the same year as the English version of Ellul's *The Technological Society*. The ideas of Ellul, Horkheimer, and Marcuse fell on fertile soil in the early 1960s. In 1964, the nuclear arms race between, mainly, the United States and the Soviet Union was accelerating. The actual and potential devastation wrought by nuclear weapons was, for many at the time, the most compelling demonstration of the dangers of out-of-control, new technology. Even two decades before atmospheric physicists first alerted us to the risk of catastrophic, "nuclear winter" scenarios, it was clear to the educated public that all-out, nuclear war between the United States and the Soviet Union could mean the end of humankind. What better proof could there be that the "technological imperative," as Ellul termed it (1964, 21), was leading us to ruin?

The early 1960s also witnessed the birth of the environmental movement. *Silent Spring*, Rachel Carson's clarion call to action about the dangers of synthetic pesticides, was published in 1962. In it, Carson marshaled evidence that DDT, in particular, was implicated in the death of many bird species, especially raptors, along with cancer and other diseases in humans exposed to such toxic chemicals (Carson 1962). Equally significant was the controversy over the construction of the Glen Canyon Dam on the Colorado River, which was started in 1956 and completed in 1966. As Lake Powell began to fill behind the dam, 186 miles of extraordinarily beautiful canyon land was flooded, with the destruction of precious habitat for rare and endangered plant and animal species as well as dozens of Native American archaeological sites. For many future, radical, environmental activists, like Edward Abbey, this was an egregious assertion of the technological domination of nature, and convinced them of the need for, sometimes, even violent resistance to the destruction of the natural environment (Abbey 1959, 1968).

While some blamed the environmental crisis on an exploding, global population, more and more thinkers were making the connection between technology and environmental problems. No one was more influential than the Washington University cell biologist and plant physiologist, Barry Commoner, whose engagement with the effects of technology on the environment began with his work in the 1950s on the environmental and health effects of radioactive fallout from atmospheric nuclear testing. By the mid-1960s, he was thinking about environmental problems from a more comprehensive point of view, as, in his 1966 book, *Science and Survival*, where he emphasized the unpredictable consequences of new technologies unleashed in a highly complex, global ecosystem and wrote that "the age of innocent faith in science and technology may be over" (Commoner 1966, 3). In his 1971 book, *The Closing Circle*, Commoner argued that it was especially the explosive growth of the synthetic petrochemicals industry after World War II that set in motion the rapid proliferation of environmental crises. Whether it is pesticides and fertilizers or detergents, synthetic textiles, and plastics, all of this new, synthetic, organic chemistry was suddenly introduced into a biosystem that had not evolved the capacity to survive such chemistry. In the chapter titled "The Technological Flaw," Commoner concluded:

> The over-all evidence seems clear. The chief reason for the environmental crisis that has engulfed the United States in recent years is the sweeping transformation of productive technology since World War II . . . . Productive technologies with intense impacts on the environment have displaced less destructive ones. The environmental crisis is the inevitable result of this counterecological pattern of growth.
>
> (Commoner 1971, 175)

This is the context in which the field of study that became technology ethics was born. A highly theoretical technology critique birthed by continental philosophers whose politics ran the gamut from Nazism to Marxism—Heidegger, Ellul, Horkheimer, and Marcuse—surfaced in the Anglophone literature at more or less exactly the moment in the early-to-mid-1960s when the anti-nuclear movement and the emerging environmental movement were pointing the finger of blame for the world's mounting problems directly at technology. In the eyes of the philosophers, the radicalized scientists, and the activists, our uncritical embrace of technology was the problem. There is irony aplenty in the intellectual alliances that emerged at this time, the community of purpose between the crypto-Fascist, Heideggerian critique of technology and the revisionist Marxist critical theorists' technology critique being the most remarkable. But more or less everyone among the parent generation of academic technology ethics agreed that technology, itself, or the socio-political embedding of technology, was to be blamed for many of the era's ever more numerous and serious woes.

We should pause to reflect on a couple of noteworthy features of the context in which technology ethics was thus born and to explore the consequences for the later development of the field. On the philosophical side, two points stand out. First, the apologists for Heidegger consistently evaded the question of the impact of his politics and his anti-Semitism on his technology critique and on his more general, philosophical project. I noted above the irony of Heidegger's seeing technology as a threat to the very essence of humankind when he had made himself an apologist for a Nazi political movement that pioneered technologized mass murder on an unprecedented scale. But the more serious worry is that both Heidegger's philosophy and his politics derived from deeply reactionary and profoundly anti-modern, anti-Enlightenment cultural roots. Long before the publication of Heiddegger's *Schwarze Hefte* [*Black Notebooks*], starting in 2014 made the connection between his politics, anti-Semitism, and philosophy undeniable (because here Heidegger makes those connections in his own words, see [Heinz and Kellerer 2016]), thoughtful readers of Heidegger already saw the connections clearly, and careful historical scholarship had laid open to view the dubious origins of his intellectual development (see Ott 1988). From his teenage years, Heidegger was shaped by an extremely conservative, south-German, Catholic world view that, even more strongly than in the official, Catholic, anti-modernist movement, repudiated the Enlightenment celebration of reason and science as the keys to human emancipation and material progress, and embraced a neo-Romantic yearning for a more authentic, pre-industrial, agrarian form of life. Heidegger's technology critique is a pure expression of this world view. What this means is that, to the extent that the birth of technology ethics as a field can be traced to Heidegger, it was not an immaculate conception.

The second point about the intellectual origins of technology ethics, even more worrisome than the questionable origins of Heidegger's technology critique, concerns the critical theorists' reification of technology or technological rationality as forces unto themselves. Careful students of Marxist dialectic should have known better than to hypostatize technology and technological rationality as something that lived apart from the material conditions of production and the rhetorical legitimation of the class interests of those who valorized technology as, of necessity, inherently a force for good. And, yet, technology and technological rationality, themselves were styled as the enemy, much as Ellul had targeted the metaphysical abstraction that he termed, "technique." These tropes were taken up in the broader community of thinkers birthing technology ethics, as illustrated by Commoner's arguing that technology was the cause of the environmental crisis.

There are two reasons why this is so worrisome. The first is that, if technology and technological rationality, themselves, are the enemy, then so, too, by implication, are the technologists, those who make technology, or, in other words, the scientists and the engineers. As a result, the assumption takes root in nascent technology ethics that, merely by virtue of one's status as a scientist or engineer, and regardless of one's self-understanding as a moral agent, one is morally implicated in the harm wrought by the technological juggernaut. The scientist and the engineer are, thereby, constructed not as allies in the effort to make a better world and promote human flourishing, but as enemies.

The second problem with the hypostatizing of technology and technological rationality as the enemy is that it steers thinking away from the choices that individual humans, corporations, government agencies, and other actors make. It also risks making malevolent technology seem to be an unstoppable force. It is a few, short steps to despair, to the crippling sense of the tragic nature of the human condition that Snow identified as key to the cultural pessimism of the literary humanist. Technological hubris is but a special case of the sin of cognitive hubris that led to the expulsion of Adam and Eve from the Garden of Eden. That trope was widely embraced by the critics of technology, as Leo Marx noted in his 1964 book, *The Machine in the Garden* (Marx 1964).

3. A Troubled Adolescence? The Maturation of Technology Ethics

A helpful indicator of the emergence of a new discipline is the launching of a journal dedicated to the scholarship that defines a community of scholarly interest. The first academic journal to become a venue of choice for literature on technology ethics appeared in 1972 in the form of the "Newsletter of the Program on Public Conceptions of Science" at Harvard University, edited by Gerald Holton. Four years later, it was taken over by Harvard's Program on Science, Technology and Human Values and given the new name, *Newsletter of Science, Technology & Human Values*. It exists today as the journal, *Science, Technology, & Human Values*, sponsored by the Society for the Social Studies of Science (see Hackett 2012). The new editor in 1976, Vivien Shelanski, wrote that the newly reconfigured journal spoke to several developments in the academy, first among them being the "surging interest in issues of scientific ethics," including the "social and scientific implications of recombinant DNA" (Shelanski 1976).

Science, Technology, & Human Values was followed in 1979 by the launch of the more specialized journal, *Environmental Ethics*, under the sponsorship of the John Muir Institute, the University of New Mexico, the American Conservation Association, and, surprising as it might seem, Chevron USA. From the start, it drew contributions from some impressive philosophers, such as Charles Hartshorne, Holmes Ralston, Michael Ruse, Tom Regan, and Mark Sagoff. Though it was by no means the journal's primary focus, a number of articles in the early years discussed the role of technology in aspects of the environmental crisis, including my own paper on "Commoner on Reductionism" (Howard 1979), Alan Drengson's paper contrasting the technocratic and deep ecology paradigms (Drengson 1980), and Kenneth Sayre's paper on "Morality, Energy, and the Environment" (Sayre 1981).

It was also in the 1970s that the first journals devoted to medical ethics were launched, such as the *Journal of Medical Ethics* in 1975 and the *Journal of Medicine and Philosophy* in 1976. Ethical challenges of biomedical technology were occasionally a focus, but the medical ethics literature of that era more commonly concerned the ethics of medical practice. And it was in 1977 that that the Institute of Electrical and Electronics Engineers (IEEE) Society on Social Implications of Technology transformed a newsletter into the magazine, *Technology and Society*. Beyond that, there were at the time no other academic journals where technology ethics literature regularly appeared.

There was, however, another, non-academic venue in which articles on technology ethics appeared. *Science for the People* was a magazine founded in 1969 by an organization named "Scientists and Engineers for Social and Political Action" (SESPA), which later changed its name to match the title of the magazine. SESPA was a heterogenous group of radical, political activists, workers, students, and university-based scientists and engineers, some of them very prominent, such as Stephen Jay Gould and Richard Lewontin. In a 2003 posting, one of the original newsletter editors, the physicist, Herbert Fox, recalled about those early members: "Most wanted to be the voice of critical consciousness from within the scientific community exposing science against the people and the dangers of the misuse of science" (Fox 2003). Opposition to the alleged misuse of science and technology by the military was a major stimulus to the formation of SESPA and the launch of the magazine, including concerns about weapons technologies used in the Vietnam War, like napalm and agent orange, or research on the development of anti-ballistic missile (ABM) technologies that were seen as a seriously destabilizing development in the context of the Cold War nuclear arms race. But, from the start, environmental problems linked to new technologies were a frequent topic, including chemical contamination of land and water by PCBs and other highly toxic substances and mountaintop removal coal mining, which had devastating environmental and, often, social consequences.

The activist orientation of SESPA and its inclusion of many university-based scientists and engineers gave it an intellectual personality very different from more purely academic work on technology and ethics in the 1950s and 1960s by philosophers like Heidegger and Marcuse. For one thing, SESPA's approach was less theoretical and more directly political. SESPA's political activism reflected the spirit of the times. But the activist orientation was also due to the presence within SESPA of a significant number of anti-revisionist Marxists, affiliated with groups like the Progressive Labor Party, a descendent of the Communist Party USA, and the Worker Student Alliance faction of Students for a Democratic Society. A key point of contention within the international socialist movement had long been whether, as revisionist social democrats argued, ideas or mere theory could play a leading role in social change. Orthodox Marxist-Leninists held that this was a form of idealism incompatible with Marx's thesis that revolutionary change emerged out of material conditions and that theory was mere superstructure. Frankfurt school social theorists like Marcuse and Horkheimer were products of the revisionist, social democratic turn in continental socialist thought in the 1920s and 1930s. Their books were read and appreciated by the activists of the 1960s and 1970s, but leftist social theory unconnected with social action was seen by far-left Marxists of the time as country club socialism.

That SESPA included so many scientists and engineers among its members was the other determining feature of its intellectual personality. Precisely because they were scientists and engineers, they knew better than anyone among the non-technical laity what were the specific threats issuing from new technologies of the post-World War II era. Their concerns were, for the most part, not abstract, theoretical, philosophical lamentations about an out-of-control, technological Golem. No. These were practical, pragmatic folk, with long experience, as during the Manhattan Project, of the ways in which the politicians and the CEOs sought to use the fruits of their genius for ends they had never imagined. They were not skeptics about science and technology, per se. They were critics of the many ways in which science and technology were being put to use for malign ends, sincerely believing that, if only the social and political circumstances were otherwise, science and technology could be put to use "for the people." Radical geek, if I might coin a phrase, was the ethos in the pages of *Science for the People*.

In addition to specialist journals, another sociological marker of the emergence of a discipline is the creation of academic programs devoted to the topic. In the same "Editor's Introduction" to the newly reconfigured journal, *Science, Technology, & Human Values*, from which I previously quoted, Shelanski listed as the second notable feature of the then current landscape "rising academic attention in STS [Science and Technology Studies], evidenced by '175 formal programs involved in some aspect of science-and-society research and/or teaching'" (Shelanski 1976). One of those programs was my undergraduate alma mater, Lyman Briggs College at Michigan State University. An interdisciplinary, residential, undergraduate, science studies college, founded in 1967 by the visionary chemist, Frederick B. Dutton, Briggs was established with the explicit goal of realizing Snow's hope that a new generation of scientists and engineers, trained also in philosophy of science, history of science, and sociology of science, would put science to use in the service of human good (Journal of Chemical Education 1996). Snow's *The Two Cultures* was required reading in the mandatory, first-year, writing and literature course, deliberately titled, "Third Culture Rhetoric."

It is noteworthy that scientists, like the chemist, Dutton, and his physics colleague, Richard Schlegel, took the lead in the founding of Briggs. It was a similar story at Edinburgh, where the Science Studies Unit, which later birthed the Edinburgh School in the sociology of scientific knowledge (SSK), otherwise known as the "strong programme" in SSK (see further discussion, later in the chapter), was established in 1966. Under the leadership of the geneticist, C. H. Waddington, and the physicist, Peter Higgs, the Committee on Providing a Broader Basis for the Science Degree was established in 1964. The committee recruited the astronomer, David Owen Edge, previously doing science education for the BBC, to move to Edinburgh, where he became the first director of the new Science Studies Unit, and it was Edge who later arranged appointments affiliated with the Science Studies Unit for Barry Barnes, David Bloor, and Steven Shapin (Williams 2016). When the Cornell University STS program was launched in 1969, it was another chemist, Franklin Long, who took the lead and became its first director (Brand 1999, Lewenstein 2016). Likewise, at Stanford University in 1971 it was the engineers, Walter Vincenti and Stephen Kline, who championed the establishment of the STS program (Stanford 1997).

An induction from all of these examples suggests that, surprising as it might seem, the intellectual orientation of many if not most first-generation STS programs was closer to that of *Science for the People* than that of the continental technology critique of Heidegger, Ellul, Horkheimer, and Marcuse. Michigan State, Cornell, and Stanford were hardly hotbeds of Marxist radical activism. But they were home to scientists and engineers who, while prizing science and technology as, potentially, forces for good, understood that the challenges of the day called forth from scientists and engineers a better understanding of the social embedding and ethical impacts of their craft. And they did not start from the premise that science and technology were the enemy. If not radical geek, they were progressive geek.

An illuminating snapshot of the landscape of technology ethics in the early 1970s, when the first-generation STS programs were being established, is afforded by the 1972 anthology, *Philosophy and Technology: Readings in the Philosophical Problems of Technology*. It was co-edited by the philosopher, Carl Mitcham, then at Berea College in Kentucky but, later, the director of the Penn State STS program, and Robert Mackey (Mitcham and Mackey 1972). The anthology was followed one year later by Mitcham and Mackey's comprehensive *Bibliography of the Philosophy of Technology* (Mitcham and Mackey 1973).[1] It was the editors' expressed intention that the 1972 anthology would create the then new field of philosophy of technology. In the preface, they write:

> As two students coming of age in the 1960s, we found ourselves living in a decade of plastic food, landscapes that resembled the printed circuits of a portable television set, and scientific toys that were rocketed into space to take possession of the moon. Like others, we were unsettled to find ourselves locked into this era of fevered affluence surrounded by profitable poverty and ever more mechanized war. As we watched the Vietnam War become an automated battlefield with American air power, stripping both children and trees of their skin, while the evening news was punctuated with advertisements for swift cars and laxatives, our minds often closed down upon our thoughts. Yet doubting, and sometimes running, we were always forced back to the same thing, more certain than before of its dominating presence. . . . It was in searching for that core of contemporary reality which could begin to make sense out of this non-spangled darkness that we discovered technology and its philosophical problems . . . .
>
> By bringing together the essays in this anthology we hope to serve a growing need of students and teachers alike, while contributing to what we consider an important event in the history of ideas–the rise of the philosophy of technology.
>
> (Mitcham and Mackey 1972, v)

Part II (of five) is devoted to "Ethical and Political Critiques." The nine papers included there cover a wide array of topics and perspectives, but, in the main, they reflect the concern about technology voiced by Mitcham and Mackey in the preface. That tone is reinforced by what the editors chose to make the concluding essay in the volume, Webster F. Hood's, "The Aristotelian Versus the Heideggerian Approach to the Problem of Technology," which the author introduces with these words:

> Martin Heidegger's reflections on the problem of technology deserve the most serious consideration. In this paper I intend both to examine some of his leading contentions on the problem and to develop an interpretation of technology based on his philosophy . . . . I shall argue that only the Heideggerian approach to technology offers any viable hope for escaping from the clutches of nihilism as it manifests itself in the guise of modern technology
>
> (Hood 1972, 347).

Quite apart from its prominent position as the final word, so to speak, in the landmark, Mitcham and Mackey anthology, Hood's essay is important also because it would appear to be the first English-language celebration of Heidegger's technology critique and, thereby, the avenue through which Anglophone philosophers of technology and first-generation technology ethicists first learned of Heidegger's views, Heidegger's original "Die Frage nach der Technik" (Heidegger 1953) having appeared in English translation only five years later (Heidegger 1977).[2] But, even if it were not the first English-language paean to Heidegger's views on technology, Hood's essay was, thanks to its inclusion in the Mitcham and Mackey anthology, the agent of the canonization of Heidegger's "The Question Concerning Technology" in the Anglophone philosophy of technology literature.

Mitcham and Mackey's goal of creating a new field of the philosophy of technology was realized in 1975 with the first of two conferences at the University of Delaware out of which would emerge in 1976 the Society for Philosophy and Technology (SPT). Mitcham, himself, assumed the role of SPT's first president in 1981 (Techné 1995). SPT launched its journal, *Techné*, in 1995 and remains, to this day, the major professional association in the philosophy of technology. Its membership in the early years reflected in their scholarship mainly the orientation of the Mitcham and Mackey anthology, though it must be said that there were dissenters from the pro-Heidegger, anti-technology orientation, an interesting case in point being the philosopher, Joseph Pitt, who established Virginia Tech's Humanities, Science, and Technology program in the mid-1970s and its Center for the Study of Science in Society in 1979. Pitt later recalled about the early years of SPT and philosophy of technology as a field: "There have been several attempts to meet the requirement of a canon, or to create something to fill the need. Such claims have been made for Heidegger's *Question Concerning Technology*, but this is at best a cult item, not a significant philosophical text." (Pitt 1995, 19). Pitt was mainly remarking on his own, very different training as an analytic philosopher who did not recognize in Heidegger his way of thinking about technology. But he went on to lament specifically the "negative" orientation of philosophy of technology in those years:

> Now, don't get me wrong. There is nothing wrong with being concerned about the adverse impact of technological developments. Nor is there anything wrong in actively being engaged in trying to avert those consequences . . . . I can be as anti-technology as anyone. But to limit your philosophical horizons to just those issues is to lose sight of what it is to be a philosopher. And for SPT to be viewed . . . as a narrowly concerned social advocacy group, is to open us up to rejection by the broader philosophical society. Our situation is no different from the Society of Christian Philosophers when they decided to make the legitimation of Christianity their agenda. We are seen as having merely a negative objective.
>
> (Pitt 1995, 20)

The dominant, anti-technology tone of the Mitcham and Mackey anthology, a collection assembled by philosophers, stands in striking contrast with the apparent commitments of the scientists and engineers who, as we have seen, played a leading role in establishing and directing the earliest STS programs. For the most part, the scientists and engineers, radicalized by the threat of nuclear weapons, the environmental crisis, and the Vietnam War, saw clearly the manifold ways in which the products of their crafts were put to damnable or, at least, morally questionable uses and resolved to do what could to make sure that such would not be the case in future. But they did not blame science and technology, per se, as the source of the problem. They blamed the political process, the policymaking process, and the decisions of individual, human actors in positions of responsibility in industry, government, the military, the media, and other loci of influence. The scientists and engineers were not naïve about such things as technological determinism or the manner in which the social, cultural, political, and economic embedding of decision makers constrained their judgment. What drove their commitment to those early STS programs was the determination not to repudiate technology, but to train future decision makers differently, with more sensitivity to and sophistication about the social and political determinants and impacts of science and technology. Mitcham, himself, noted this tension in his 1994 book, *Thinking through Technology*, where he distinguished the "engineering tradition" and the "humanities tradition" (Mitcham 1994, Ihde 1995, 9). As Snow had argued more than a decade earlier, it was radical geek versus dour pessimism about technology's role in human affairs.

As the first-generation STS programs were getting off the ground and the literatures that would come to define philosophy of technology and technology ethics were being collected and canonized by Mitcham and Mackey, other intellectual currents were stirring in the academy. While it took a few years for the full impact to be felt, the transformative events of 1968—the surging anti-war movement in the United States, the Paris riots, the Soviet invasion of Czechoslovakia—set off a transformation in philosophy. The legacy of McCarthyism and the "Red scare" in the United States was waning. Radical movements were sweeping through Western Europe and North America. The post-World War II, liberal democratic status quo was under assault everywhere. One particularly important expression of the changing times was the re-emergence in Germany of the tradition of Frankfurt School critical theory in the person of Jürgen Habermas, whose hugely influential book, *Erkenntnis und Interesse* [*Knowledge and Human Interests*], appeared in 1968 (Habermas 1968a), its English translation following in 1971 (Habermas 1971).

Habermas revived Horkheimer's project of the 1930s, expanding upon Horkheimer's critique of "traditional theory" in the human and the natural sciences. Both are faulted for their inability to support genuinely emancipatory and transformative, critical theoretical reflection on human life, society, politics, and economic relations. "A radical critique of knowledge is possible," Habermas wrote, "only as social theory" (Habermas 1971, vii). He argued that philosophy of science in the form of positivism epitomizes the problem by emphasizing description, prediction, and control instead of the kind of critical reflection that can subvert received self-understandings and structures of power and authority. Technology was not the main focus of Habermas's argument, but, rather, science as theorized within a positivist framework that prizes prediction and control for the purpose of technological innovation, with the implication that human emancipation and flourishing comes mainly in the form of growing material well-being. Technology, or rather the rise of what he terms "technocratic consciousness," was, however, the focus of an essay that Habermas wrote in 1968 in honor of Marcuse's seventieth birthday, "Technik und Wissenschaft als 'Ideologie'" ["Technology and Science as 'Ideology'"] (Habermas 1968b). Habermas argued here that the "scientization of technology," meaning the intentional and systematic application of science in planned technological innovation, had transformed the nature and role of technology with unfortunate consequences for emancipatory political projects:

> Technocratic consciousness reflects not the sundering of an ethical situation but the repression of "ethics" as such as a category of life. The common, positivist way of thinking renders inert the frame of reference of interaction in ordinary language, in

> which domination and ideology both arise under conditions of distorted communication and can be reflectively detected and broken down. The depoliticization of the mass of the population, which is legitimated through technocratic consciousness, is at the same time men's self-objectification in categories equally of both purposive-rational action and adaptive behavior. The reified models of the sciences migrate into the socio-cultural life-world and gain objective power over the latter's self-understanding. The ideological nucleus of this consciousness is the elimination of the distinction between the practical and the technical. It reflects, but does not objectively account for, the new constellation of a disempowered institutional framework and systems of purposive-rational action that have taken on a life of their own.
>
> (Habermas 1968b, 112)

This is not the place to quibble about Habermas's caricature of twentieth-century philosophy of science. The point to emphasize here is that Habermas's revival and reinvigoration of Frankfurt School critical theory introduced a new, post-1968 generation to the science and technology critique pioneered by Marcuse and Horkheimer three decades earlier.

Habermas's emergence as a leading voice in the 1970s literature on the political and cultural impact of science and technology coincides roughly with another development of some moment for philosophical and social critiques of science and technology, this being the rapid rise to prominence of the aforementioned, self-styled "strong programme" in the sociology of scientific knowledge (SSK), the "Edinburgh School." Chiefly the work of the Barry Barnes (Barnes 1974, 1977) and David Bloor (Bloor 1976), the strong programme distinguished itself from the work of earlier sociologists of science, such as Robert K. Merton, with the contention that social and political context could shape not only the institutional structures of science, including such things as the setting of research agendas, but also the very content of scientific theories. That was the sense in which it was a "strong" program. This was hardly a new idea, but it did constitute something of a revolution in science studies in the 1970s and beyond, and it afforded scholars new tools from sociology and anthropology for analyzing the social, political, and economic embedding of science. By intention, it also problematized unreflected claims to scientific objectivity and, thereby, to the cultural authority of science. As with the work of Habermas, technology was not the main target of strong programme SSK, but, to the extent that it called into question the cultural authority of science, so, too, it called into question the cultural authority of scientifically driven, technological innovation and the associated institutional structures for social control through technology.

4. Are We All Adults Now? Technology Ethics Matures

Such was the birth and the early history of the field of technology ethics. More than forty years have passed since the mid-1970s, where my historical narrative stops. If I might be permitted a broad and contentious generalization, not much happened in technology ethics in those intervening forty years by way of changing the dominant orientation of the literature. To be sure, the problems that we debate today are markedly different. Anthropogenic climate change was not at all well understood in the mid-1970s. The full import of the microelectronics and internet revolution was beyond our imaginative powers. Even John von Neumann and Alan Turing had not foreseen the transformative impact of artificial intelligence and machine learning in areas as diverse as medical diagnostics, automated trading platforms, and autonomous weapons. No one had any sense of how different our world would be thanks to the technical infrastructure supporting modern social media. And that we might someday easily and cheaply manipulate human, other mammalian, plant, and microbial genomes at the level of individual base pairs was only the stuff of science fiction, not serious, technological prognostication. All of these challenges have called forth sophisticated and incisive scholarship. The expanding literature in contemporary technology ethics is as varied as is the problematic landscape, and the technical quality of that literature improves steadily. But the basic polarity of this scholarship remains as it was set in the very specific cultural, political, and philosophical circumstances of the field's birth, with the philosophers, historians, and social theorists still, for the most part, playing their assigned role as Snow's cultural pessimists. To be socialized into the technology ethics community from the philosophical side is to be trained to look mainly for the risks and dangers accompanying new technologies and to assume the part of either the alarmist or the counselor of prudence and precaution.

That this monitory orientation still dominates the technology ethics literature is what should have been expected given the history of the field's birth and adolescence as sketched in this paper. But history need not be fate, and one reason for studying history is to overcome it. If technology ethics is to mature into a professional discipline that effectively engages the rapidly proliferating challenges presented by technological innovation that has passed the inflection point in its exponential acceleration, if it is to be more than mere academic discourse in an echo chamber of the like-minded, if it is to be heeded by policymakers, corporate executives, engineers, and consumers, then it must grow beyond the sureties and enthusiasms of its youth and learn to practice also an ameliorative form of scholarship, one that seeks, finds, and promotes technological innovation that empowers, emancipates, and enhances human well-being. That ameliorative project requires as much philosophical sophistication, historical knowledge, and analytical insight as is needed for the discernment of risk, and it might well require considerably more by way of creative, philosophical imagination. Understanding and explaining the ways in which technological innovation can promote human flourishing is as much a philosophical task as are any philosophical reflections on the nature of the good life and the various paths to its realization. But an ameliorative technology ethics cannot even get off the ground if we begin with the premise that technology is the daemon driving the tragedy of human existence in the modern era.

The call for a more technology-friendly technology ethics is sometimes met with the assertion that it is unnecessary because there is no shortage of technology promoters in industry, government, and the media. But this objection misses the point. Of course corporations promote their new products with recitals of all of the wonderful benefits that will supposedly flow from one's owning the newest smartphone or a home thermostat that one can control from one's office laptop. Marketing hype is, however, rather a different thing from careful philosophical analysis and argumentation. Moreover, market-driven technology innovation seeks profit first and the betterment of the human condition only as an afterthought. An amelioratory technology ethics, by contrast, starts with a vision of the good life and then asks how technology can help to attain it. An amelioratory technology ethics also differs from more hype and propaganda by being always alert to risk. Prudence is essential to the pursuit of the good life.

Let me conclude with two examples that illustrate an amelioratory complement to monitory technology ethics. The first, and probably less contentious example, concerns self-driving vehicles (SDVs). There is, already, a large literature on ethical problems with SDVs (see, for example, the papers on autonomous vehicles in Lin, Jenkins, and Abney 2017), and some of the more widely discussed issues, such as the trolley problem as applied in the case of SDVs, have made their way from the scholarly literature into the popular press (see, for example, Lin 2013). As with so many other issues in technology ethics, the dominant tone here is monitory. The philosophers are pointing out important ethical concerns, from puzzles about how to program SDVs to make morally fraught decisions when collisions are unavoidable and the dilemma of moral responsibility attribution when an SDV is not controlled by a human agent, to the big problem of technological unemployment when SDVs throw millions of truck and taxi drivers out of work.

More or less entirely missing from the philosophical literature on the ethics of SDVs are, however, discussions of the moral gains promised by the adoption of this technology, from enhancing the autonomy and flourishing of people with disabilities that previously made difficult or impossible their employment of personalized transportation, to reducing drastically the number of deaths and injuries from vehicle accidents because of the inherent superiority of autonomous control systems over human drivers. Missing as well is philosophical reflection on strategies for promoting the responsible but rapid deployment of a technology that promises so much moral gain through regulatory reform, market incentivization, and the reconstruction of infrastructure. Philosophers trained in skills of analysis and persuasion are well suited to help solve such problems. There are few other opportunities for applied ethicists to help save the lives of 1.2 million people annually.

The other example, sure to be a far more contentious one, concerns nuclear power generation. The accidents at Three Mile Island (1979), Chernobyl (1986), and Fukushima (2011) reinforced long-standing fears about the risks of nuclear power and joined other famous environmental catastrophes as iconic representations of the dangers of technology run amok. The first two also loomed large in shaping the thinking of philosophers, legal scholars, scientists, and engineers about the importance of the precautionary principle in technology innovation (see Steel 2015). Philosophers have long been involved in debates about the ethics of nuclear energy, but the Fukushima accident led to a renewal of interest in the topic, as witnessed by the recent collection, *The Ethics of Nuclear Energy: Risk, Justice, and Democracy in the Post-Fukushima Era* (Taebi and

Roeser 2015). A variety of views are aired here, but, as is typical of the field of technology ethics, the dominant orientation is monitory, with the emphasis on environmental risk and questions of social justice. One is right to be concerned, of course. In the end, little harm was done by the Three Mile Island accident, but the Chernobyl accident caused at least several score deaths, principally among first responders, and mass evacuations led to long-term, social and economic disruptions for tens of thousands, as did the mass evacuations at Fukushima. Of equal or greater concern is the problem of nuclear waste disposal, which poses major technical challenges.

Lacking, however, in most of the philosophical literature on nuclear energy is the requisite comparative ethical perspective. The question is not, simply, whether to generate electricity from burning nuclear fuel. The question is about the ethical impact of nuclear energy in comparison with other forms of energy production (see Howard 2020). Such a comparative perspective would yield important conclusions in two areas. First, as concerning as are the Three Mile Island, Chernobyl, and Fukushima accidents, the human suffering that they caused pales in comparison to the human toll from generating electricity by burning coal, oil, and natural gas, but, especially, coal. Included in that grim calculus are the hundreds of thousands of deaths in coal mining over two centuries, deaths caused by accidents and the health effects of long-term exposure to coal dust, such as brown lung. Included, as well, must be the millions of early deaths and the millions of cases of non-lethal disease from breathing air polluted by burning coal and drinking water contaminated by run-off from coal piles and ash pits. Add to this the other, massive, environmental impacts of coal mining and the burning of coal, from decapitated mountains to the acidification of lakes, streams, and oceans. There is also that little problem of anthropogenic climate change, which is mainly a result of burning fossil fuels for energy production.

A second consequence of the adoption of a comparative perspective is the realization that, among the "green" alternatives for energy production, nuclear is the only one that is scalable and that can address the "base load" problem, which is the need to produce electricity on demand, at any hour of the day or night, regardless of cloud cover and wind speed, anywhere in the world. One would expect that technology ethicists would lead the way in counseling us that questions of moral choice are always questions about choice among options, hence comparative questions, and that, almost never, will there be a morally perfect choice, so that, again, it is not a question of whether moral cost attaches to nuclear energy, considered by itself, but a question of which choice, among several imperfect options, maximizes the likelihood of human flourishing. An amelioratory complement to a monitory ethics of nuclear energy technology would emphasize that insight.

5. Conclusion

The circumstances of its birth inclined the still developing field of technology ethics toward a mainly monitory orientation, emphasizing attention to risk and caution. But for the field really to have the impact that it must outside of the academy, for it to engage constructively with technologists, policymakers, and consumers, that risk-averse orientation should be complemented by an amelioratory orientation, one that is equally alert to opportunities for technological innovation and deployment that can enhance human flourishing. This should not be confused with what some deride as "technosolutionism," the naïve idea that technology, alone, can be our salvation. Of course social and political interventions and innovations are critical, and an amelioratory technology ethics must always evaluate specific technologies in the sociocultural contexts in which they are and will be employed. The main point is, rather, that the field must balance despair with hope and timidity with courage.

Acknowledgments

I want to thank Shannon Vallor for the invitation to contribute to this Oxford Handbook, for her support and encouragement when unexpected circumstances made the completion of the paper more challenging that it otherwise should have been, and her patience in waiting for the final product. This paper also benefitted from the input of many friends and colleagues, among them Carl Mitcham, Patrick Gamez, Andrea Hoplight Tapia, Bruce Lewenstein, David Kaiser, Trevor Pinch, Anna Geltzer, Matt Brown, Nicholas Rowland, Gary Weisel, Evan Selinger, Darrin Durant, George Gale, Elam O'Renick, Hasok Chang, David Guston, Alexei Kojevnikov, and Joseph Pitt.

Notes

[1] Equally important in creating the new field of philosophy of technology were the annual compilations of research in the area compiled by Mitcham and Paul Durbin (Durbin and Mitcham 1978–1985).

[2] Mitcham and Mackey sought to do an English translation of Heidegger's "Die Frage nach der Technik" (Heidegger 1953) for publication in the 1972 anthology, but they were refused permission by the German publisher of Heidegger's *Vorträge und Aufsätze*, Gunther Neske (Carl Mitcham, private communication.)

References

Abbey, Edward. 1959. "Anarchism and the Morality of Violence." Master's Thesis. University of New Mexico.

Abbey, Edward. 1968. *Desert Solitaire: A Season in the Wilderness*. New York: McGraw-Hill.

Barnes, Barry. 1974. *Scientific Knowledge and Sociological Theory*. London and Boston: Routledge & Kegan Paul.

Barnes, Barry. 1977. *Interests and the Growth of Knowledge*. London and Boston: Routledge & Kegan Paul.

Bloor, David. 1976. *Knowledge and Social Imagery*. London and Boston: Routledge & Kegan Paul.

Brand, David. 1999. “Franklin A. Long Dies at 88; Was Cornell Professor and Controversial Figure in Nixon Era.” *Cornell Chronicle*, February 10, 1999. https://news.cornell.edu/stories/1999/02/franklin-long-dies-88-was-cornell-professor-and-controversial-figure-nixon-era.

Carson, Rachel. 1962. *Silent Spring*. Boston: Houghton Mifflin.

Commoner, Barry. 1966. *Science and Survival*. New York: Viking Press.

Commoner, Barry. 1971. *The Closing Circle: Nature, Man, and Technology*. New York: Knopf.

Dessauer, Friedrich. 1927. *Philosophie der Technik. Das Problem der Realisierung*. Bonn: Friedrich Cohen.

Drengson, Alan R. 1980. “Shifting Paradigms: from the Technocratic to the Person-Planetary.” *Environmental Ethics* 2: 221–240.

Durbin, Paul T., and Carl Mitcham, eds. 1978–1985. *Research in Philosophy & Technology*. Greenwich, CT: JAI Press.

Ellul, Jacques. 1954. *La Technique: L’Enjeu du siècle* Paris, Armand Colin.

Ellul, Jacques. 1964. *The Technological Society*. John Wilkinson, trans. New York: Knopf. (Translation of Ellul 1954.)

Fox, Herbert. 2003. Science for the People Listserv, May 19, 2003. https://web.archive.org/web/20070928192107/http://www.scienceforthepeople.com/modules.php?op=modload&name=News&file=article&sid=21.

Habermas, Jürgen. 1968a. *Erkenntnis und Interesse*. Frankfurt am Main: Suhrkamp.

Habermas, Jürgen. 1968b. “Technik und Wissenschaft als ‘Ideologie.’ ” Merkur, 591–610, 682–693. Reprinted in *Technick und Wissenschaft als ‘Idologie.’* Frankfurt am Main: Suhrkamp, 48–103. Page numbers and quotations from the English translation, Jürgen Habermas, 1970. “Technology and Science as ‘Ideology.’ ” In *Toward a Rational Society: Student Protest, Science, and Politics*, translated by Jeremy J. Shapiro, 81–121. Boston: Beacon Press, 1970.

Habermas, Jürgen. 1971. *Knowledge and Human Interests*, translated by Jeremy J. Shapiro. Boston: Beacon Press.

Hackett, Edward J. 2012. “Science, Technology, & Human Values at 40.” *Science, Technology, & Human Values* 37: 439–442.

Halberstam, David. 1972. *The Best and the Brightest*. New York: Random House.

Heidegger, Martin. 1953. “Die Frage nach der Technik.” Lecture on November 18, 1953 at the Technische Hochschule, Munich, in the series, “Die Künste im technischen Zeitalter” [“The Arts in the Age of Technology”], under the auspices of the Bavarian Academy of Fine Arts. Bayerische Akademie der schöne Künste. *Jahrbuch*. Munich: R. Oldenbourg, 1954, 70ff. Reprinted in Martin Heidegger. 1954. *Vorträge und Aufsätze*. Pfullingen, Germany: Günther Neske, 13–44. Page numbers and quotations from the English translation, Heidegger 1977.

Heidegger, Martin. 1977. “The Question Concerning Technology.” In Martin Heiddeger. *The Question Concerning Technology and Other Essays*, translated by William Lovitt, 3–35. New York and London: Garland, 1977.

Heinz, Marion, and Sidonie Kellerer, eds. 2016. *Martin Heideggers »Schwarze Hefte«—Eine philosophisch-politische Debatte*. Berlin: Suhrkamp.

Hood, Webster F. 1972. “The Aristotelian Versus the Heideggerian Approach to the Problem of Technology.” In Mitcham and Mackey 1972, 347–363. From: “A Heideggerian Approach to the Problem of Technology.” Ph.D. Dissertation. Pennsylvania State University, 1968.

Horkheimer, Max. 1937. “Traditionelle und kritische Theorie.” *Zeitschrift für Sozialforschung* 6: 245–294. Reprinted in: Max Horkheimer. *Kritische Theorie*. Frankfurt am Main: S. Fischer. English translation Matthew J. O’Connell et al.: Max Horkheimer. 1972. *Critical Theory*. New York: Herder and Herder.

Horkheimer, Max. 1947. *Eclipse of Reason*. New York: Oxford University Press.

Howard, Don. 1979. “Commoner on Reductionism.” *Environmental Ethics* 1, 159–176.

Howard, Don. 2020. “The Moral Imperative of Green Nuclear Energy Production.” *Notre Dame Journal on Emerging Technologies*. 1: 64–91. https://ndlsjet.com/the-moral-imperative-of-green-nuclear-energy-production-2/

Ihde, Don. 1995. “Philosophy of Technology, 1975–1985.” *Techné: Research in Philosophy and Technology* 1: 8–12.

Journal of Chemical Education. 1996. “In Memoriam–Frederick B. Dutton.” *Journal of Chemical Education* 73: 107.

Leavis, Frank Raymond. 1963. *The Two Cultures? The Significance of C. P. Snow*. New York: Pantheon Books.

Lewenstein, Bruce. 2016. “Remarks at S&TS Open House and 25th Birthday Celebration.” Unpublished manuscript.

Lin, Patrick. 2013. “The Ethics of Autonomous Cars: Sometimes Good Judgment Can Compel Us to Act Illegally. Should a Self-driving Vehicle Get to Make that Same Decision?” *The Atlantic*, October 8, 2013. https://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/.

Lin, Patrick, Ryan Jenkins, and Keith Abney, eds. 2017. *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. New York: Oxford University Press.

Marcuse, Herbert. 1941. “Some Social Implications of Modern Technology.” *Studies in Philosophy and Social Science* 9: 414–439.

Marcuse, Herbert. 1964. *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*. Boston: Beacon Press.

Marx, Leo. 1964. *The Machine in the Garden: Technology and the Pastoral Ideal in America*. New York: Oxford University Press.

Mitcham, Carl. 1994. *Thinking through Technology: The Path between Engineering and Philosophy*. Chicago: University of Chicago Press.

Mitcham, Carl, and Robert Mackey, eds. 1972. *Philosophy and Technology: Readings in the Philosophical Problems of Technology*. New York: The Free Press.

Mitcham, Carl and Robert Mackey, eds. 1973. *Bibliography of the Philosophy of Technology*. Chicago: University of Chicago Press.

Mumford, Lewis. 1934. *Technics and Civilization*. New York: Harcourt, Brace & Company.

Mumford, Lewis. 1967–1970. *The Myth of the Machine*, 2 vols. New York: Harcourt, Brace & World.

Ott, Hugo. 1988. *Martin Heidegger: Unterwegs zu seiner Biographie*. Frankfurt am Main: Campus. English trans.: *Martin Heidegger: A Political Biography*. New York: Basic Books, 1993.

Pitt, Joseph C. 1995. “On the Philosophy of Technology, Past and Future.” *Techné: Research in Philosophy and Technology* 1: 18–22.

Sayre, Kenneth. 1981. “Morality, Energy, and the Environment.” *Environmental Ethics* 3: 5–18.

Shelanski, Vivien B. 1976. “Editor’s Introduction.” *Newsletter of Science, Technology & Human Values*, no. 17:, 1–2.

Snow, Charles Percy. 1959. *The Two Cultures and the Scientific Revolution*. Cambridge: Cambridge University Press.

Snow, Charles Percy. 1963. *The Two Cultures: and A Second Look*. Cambridge: Cambridge University Press.

Stanford University. 1997. "Memorial for Stephen Kline; Engineer, Interdisciplinary Thinker." Stanford University News Release. https://news.stanford.edu/pr/97/971028kline.html.

Steel, Daniel. 2015. *Philosophy and the Precautionary Principle: Science, Evidence, and Environmental Policy*. Cambridge: Cambridge University Press.

Taebi, Behnam, and Sabine Roeser. 2015. *The Ethics of Nuclear Energy: Risk, Justice, and Democracy in the Post-Fukushima Era*. Cambridge: Cambridge University Press.

Techné. 1995. "Introduction to the New Journal." *Techné: Research in Philosophy and Technology* 1, 2.

Williams, Robin. 2016. "Still Growing Strong. 50 Years of Science, Technology, and Innovation Studies at the University of Edinburgh." *EASST Review* 35 (3). https://easst.net/article/still-growing-strong-50-years-of-science-technology-and-innovation-studies-at-the-university-of-edinburgh/.

PART II

TECHNOLOGY AND EPISTEMOLOGY

CHAPTER 6

STYLES OF OBJECTIVITY IN SCIENTIFIC INSTRUMENTATION

A. S. AURORA HOEL

1. Introduction

This chapter is situated at the intersection of two disciplines: the philosophy of technology and science studies (taken broadly, including the history and philosophy of science). It contributes to the current attempts to conceptualize the productive roles of technologies in knowledge formation, with a special focus on scientific instruments. The chapter draws on the example of magnetic resonance imaging (MRI), which has become vitally important in day-to-day clinical practice in hospitals across the world—including in neuroradiology, where MRI is currently the go-to tool in the diagnosis of tumors and other lesions of the brain.

My point of departure is that technologies have *agency*. This is an assumption that, in recent years, has come to be shared by a great many scholars of a broad range of different disciplines investigating the epistemic roles of technology. But what is this agency, and what are the philosophical implications of acknowledging it? Various concepts and metaphors have been proposed to express the growing realization that agency is not exclusively human. Examples include "mangle of practice" (Pickering 1995) and "entanglement" (Latour 2005), both of which emphasize the distributed and interdependent nature of agency. Other metaphors emphasize how humans and technologies mutually shape each other, scholars talking about the "co-shaping" or "co-constitutive" roles of technologies (Verbeek 2005). This chapter adds to the growing vocabulary by approaching technologies as what I call "adaptive mediators." While the notion of adaptive mediator retains the focus on distributed and interdependent agencies, including an overall relational outlook, it differs from leading science-studies approaches by pushing further into an ecological and operational conceptual terrain.

The title of this chapter may invite readers to think of other uses of the style notion in connection with scientific knowledge formation. A prominent example is Ian Hacking, who coined the term "styles of reasoning" (Hacking 1982, 1992, and 2012). A notable historical example is Ludwik Fleck, who talked about "thought style" and "thought collective" (Fleck 1979). The approach suggested in this chapter resonates with Hacking's and Fleck's in that it emphasizes the formation of distinct modes of inquiring into nature and ourselves, which are often formalized into shared "ways of finding out in the sciences," as Hacking puts it (Hacking 2012, 601). These examples further resonate with a broader family of analytical notions, including Thomas Kuhn's "paradigm" (Kuhn 1962) and Michel Foucault's "discourse" or "episteme" (Foucault 1970). Each in their own way, these approaches carry on the Kantian project of explaining how objectivity in science is possible, with the crucial exception that they all set out to *historicize* the Kantian categories. Again in the words of Hacking: "Kant did not think of scientific reason as a historical and collective product. We do" (Hacking 1992, 4). The historicizing move has the effect of replacing the Kantian table of categories with "a sort of historical a priori" (Foucault 1970, 172). But the historicizing move comes at a price: The historical canons of objectivity lack the necessity that was essential to the Kantian pure concepts of the understanding.

The approach to be sketched in this chapter aligns with the approaches mentioned in that it probes into the conditions of knowledge and performs a historicizing move. Yet it differs in that it proceeds to perform a second move, which I call "ecologicizing." As we shall see, the ecologicizing move has the effect of challenging the cognitive dualism at the heart of the Kantian model: the dichotomy between a receptive faculty of sensibility and an active faculty of understanding. What is more, the ecologicizing move has the effect of putting technical mediation at the center of epistemology, shifting away from the prevailing tendency of theories of knowledge to focus primarily (and in many cases exclusively) on thought and language.

Since the end of the 1970s, the social studies of science have been, as Hacking notes, a hub of innovation in the philosophy of science (Hacking 1999a, 186). That said, the new developments have also spurred much controversy. The controversy turns on the social constructionist assumptions underpinning much science studies research, having given rise to heartfelt disagreements commonly referred to as "the science wars": bitter debates where scholars lumped under the label "scientific realists" clash with scholars lumped as "postmodernists" (Hacking 1999a, vii; Ihde 2009, 5). Thus, while social construction has inspired a plethora of trailblazing empirical studies of science-in-the-making (Latour and Woolgar 1979, Pickering 1984), it has also triggered much anger, which Hacking relates to "a great fear of relativism" (Hacking 1999a, 4).

On a more cheerful note, during the same time period, there has been a rapprochement between the philosophy of technology and the social studies of science. As pointed out by Hans Achterhuis, over the last decades, the philosophy of technology has undergone "an empirical turn that might roughly be characterized as constructivist" (Achterhuis 2001, 6). In this context, the label "empirical turn" implies a move away from the concerns of the classical philosophers of technology, who, according to Achterhuis, “occupied themselves more with the historical and transcendental conditions that made modern technology possible than with the real changes accompanying the development of a technological culture” (Achterhuis 2001, 3). In presenting his own approach to technology called “postphenomenology,” Don Ihde agrees with Achterhuis’ characterization of the empirical turn and proceeds to maintain that postphenomenology “is a step into the style of much ‘science studies,’ which deals with case studies” (Ihde 2009, 22).

The latter comment points to the ongoing reconciliation between philosophy of technology and the social studies of science, especially STS. To this end, efforts have been made by postphenomenologists to compare the perspective of postphenomenology with that of actor-network theory (ANT). In most cases, these comparisons have been guided by the conviction that the two methods of analysis “are more complementary than combative” (Ihde 2015, xvi). For all that, in situations where postphenomenologists are forced to explicate what sets postphenomenology apart, several important differences tend to be noted. In the words of Ihde, while both styles of analysis “are materially sensitive” and “employ inter-relational ontologies,” they differ in that ANT “draws from semiotics of which the base is linguistic-textual,” whereas postphenomenology “draws from an embodiment analysis of human action and perception” (Ihde 2015, xv). He also encapsulates what distinguishes postphenomenology by characterizing it as a method of analysis that is “closer to an ‘organism/environment’ model than is often appreciated” (Ihde 2003, 133).

The latter remark by Ihde is the key to the rest of this chapter. The overarching aim of the chapter is to investigate the philosophical implications of replacing the familiar subject/object model with an organism/environment model. It starts out by exploring the organism/environment model in some detail, showing how a developed version of this model gives rise to a new notion of relationality, which I call “ecological relationality”—a notion that differs in philosophically significant respects from the “poststructuralist relationality” (Law 2009, 145) that underpins ANT. It proceeds to examine some of the epistemological implications of ecological relationality, pointing to how the ecological model opens the way for new epistemologies beyond the deadlocked positions of the science wars.

However, for the organism/environment model to do all this, it needs to be developed in a specific direction, namely, along a conceptual trajectory that emphasizes the *operational* aspects of technical mediation. To arrive at the extra steps that are required for the ecological model to unlock new epistemologies, I have been helped along by new developments in contemporary media theory. These developments offer new inroads into the study of mediation by emphasizing the operational aspects of technologies and media (Farocki 2004), including their role in providing *conditions* for our lived environments (Mitchell and Hansen 2010, Peters 2015). However, when it comes to developing the pivotal notion of adaptive mediator, I draw on historical sources, more precisely on the thinking of Gilbert Simondon—a French philosopher whose remarkably original ideas about technology are currently being rediscovered in the philosophy of technology, science studies, media theory, and beyond.

2. Foundational and Relational Ecological Models

Michel Foucault's "episteme" and Ian Hacking's "styles of reasoning" are both rooted in a French philosophical tradition that includes thinkers such as Gaston Bachelard and Georges Canguilhem, who were already underway to historicize epistemology (Sciortino 2017, 257). In his famous introduction to the English translation of Canguilhem's *The Normal and the Pathological*, Foucault identified two opposing traditions in postwar French philosophy: "a philosophy of experience, of sense and of subject" espoused by thinkers like Jean-Paul Sartre and Maurice Merleau-Ponty, and "a philosophy of knowledge, of rationality and of concept" espoused by thinkers like Bachelard and Canguilhem (Foucault 1991, 8). In distinguishing the two traditions, Foucault positioned himself as belonging to "the conceptualist side," distancing himself from "the subjectivist side" (Rheinberger 2005, 313).

Even though Foucault's division is questionable on several counts (a point I will return to), it is worthwhile to take a closer look at his reasons for making it. This is because Foucault's qualms about putting too much emphasis on sensibility and the body are indicative of another fear that resurfaces in the science wars, which I will call "the great fear of foundationalism." The consideration of Foucault's qualms also allows me to make two important clarifications about the model proposed in this chapter: that it is a *relational* rather than foundational model, and that it *no longer adheres to the dichotomy of life and thought*, which Foucault tacitly assumes when making his division.

2.1 Bachelard and the Notion of Phenomenotechnique

Bachelard is a key figure in the "conceptualist" tradition. He introduced many of the notions that were later popularized by Foucault (and others), such as "epistemological break" and "epistemological obstacle" (Gutting 1989, 52). For Bachelard, the notion of epistemological break relates to discontinuities in the development of science, including how previous scientific conceptions can become obstacles in the pursuit of truth, hampering the formation of the "true scientific mind" (Bachelard 2002, 25). Crucially, however, the notion also relates to how the scientific mind breaks away from "primary experience," which presents an even more serious obstacle to truth (2002, 20, 33). "Science," as conceived by Bachelard, "is totally opposed to opinion" (2002, 25), whether in the form of old prejudices or everyday customs. The true scientific mind, therefore, must be constituted against false science on the one hand, and against nature (in the form of primary experience) on the other (2002, 33, 38). To account for the objectivity of scientific knowledge and to explain what it means for the true scientific mind to be formed "*against* nature," Bachelard introduces the term "phenomenotechnique." The hallmark of modern experimental science is that it "*realises* its objects without ever just finding them ready-made" (2002, 70, original emphasis). This means that a concept becomes scientific only insofar as it is accompanied "by a technique that realises," that is, by some kind of scientific instrument that "*extends* phenomenology" (2002, 70, original emphasis). Hence, for Bachelard, a truly scientific phenomenology *is* a phenomenotechnique, whose purpose is "to amplify what is revealed beyond appearance" (Bachelard 1984, 13).

Bachelard's approach is noteworthy for its insistence on the indispensability of instruments in science, and also, for the way it assigns to these instruments a productive—indeed, realizing—role. Even so, in the remainder of this chapter, I shall leave Bachelard behind, seeking instead to arrive at the productive role of instruments via a different route. I do this because Bachelard's approach is firmly based on a stark opposition between life and thought, and hence, on a cognitive dualism akin to Kant's dichotomy between sensibility and understanding. I turn instead to a selection of approaches that set out to challenge such cognitive dualisms, including the long-standing assumption that there is something essentially irrational about life and sensibility.

2.2 The Ecological Motif in Canguilhem and Merleau-Ponty

To find an example of such an approach, we do not have to look far. A prominent example is found at the heart of the "conceptualist" tradition—in fact, in the very work for which Foucault wrote his introduction. For while the ecological motif is missing in Bachelard, it is highly pronounced in Canguilhem. *The Normal and the Pathological* is widely celebrated as a major contribution to the history and philosophy of science. In this work, Canguilhem defines normality in medicine and biology in terms of the capacity of living beings to institute new "norms of life" in relation to their environments (Canguilhem 1991, 144). In arriving at this definition, Canguilhem makes extensive use of ideas developed by the German neurologist and psychiatrist Kurt Goldstein, who in turn had adopted and critically adjusted the idea of the complementarity and reciprocity of the organism and its *Umwelt*, as developed by the Estonian-German biologist Jakob von Uexküll. The idea that life has *norms*, indicates that life "cannot be the blind and stupid mechanical force that one likes to imagine when one contrasts it to thought" (Canguilhem 2008, xviii). As should be clear from this statement, Canguilhem rejects the existence of "a fundamental conflict between knowledge and life" (2008, xvii). Instead, he suggests that we approach knowledge as "a general method for the direct or indirect resolution of tensions between man and milieu" (2008, xviii).

The ecological motif is also highly pronounced in the work of Merleau-Ponty, whom Foucault, as we have seen, classified as belonging to the "subjectivist" tradition. As is the case with Canguilhem, Merleau-Ponty's major works are peppered with references to

Goldstein. In his own preface to the second edition of *The Normal and the Pathological*, Canguilhem comments on the kinship between the two approaches, regretting that he did not know the contents of Merleau-Ponty's *The Structure of Behavior* (Merleau-Ponty 1963) at the time he was developing the central theme of his own book (Canguilhem 1991, 29). Both thinkers, it seems, evoked Goldstein to help rethink the philosophical status of living beings: Canguilhem to vindicate the rationality of life, Merleau-Ponty to vindicate the rationality of sensibility. Why, then, does Foucault place the two thinkers on opposite sides of what he saw as the central cleavage in mid-twentieth-century French philosophy?

2.3 Foundational Ecological Models

A clue to the answer is found in Foucault's *The Order of Things* (Foucault 1970). In this work, Foucault observes that Kant's definition of man as "an empirico-transcendental doublet" gave rise to two kinds of analysis: There were analyses that, by studying the sensorial mechanisms of the body, discovered that knowledge has "anatomo-physiological conditions," and hence, that there is "a *nature* of human knowledge" (Foucault 1970, 319, original emphasis). On the other hand, there were analyses that emphasized that knowledge has "historical, social, or economic conditions," that it is "formed within the relations that are woven between men," and hence, that there is "a *history* of human knowledge" (Foucault 1970, 319, original emphasis). He further observes that the two kinds of analysis correspond to a more fundamental division of truth itself: The first kind of analysis is related to a truth "of the positivist type," which is "of the same order as the object" and "expressed through the body and the rudiments of perception" (Foucault 1970, 320). The second kind of analysis is related, rather, to a truth "of the eschatological type," which is "of the order of discourse," and which "anticipates the truth whose nature and history it defines" (Foucault 1970, 320). Foucault proceeds from this to identify phenomenology with positivism, claiming that phenomenology has never fully succeeded in exorcizing its "insidious kinship" to empirical analyses of man, and that "the analysis of actual experience" is nothing but a more careful attempt to make "the empirical [. . .] stand for the transcendental" (Foucault 1970, 321, 326).

Commenting on *The Order of Things*, Canguilhem notes that it would have been worthwhile for Foucault to have dealt in more detail with the case of Auguste Comte, the acknowledged father of positivism. Comte's project was to substitute "the scientific relation between organism and environment for the metaphysical relation between subject and object" (Canguilhem 1994, 87). However, since for Comte, the physiological a priori is more fundamental than the historical a priori, he proposed to found a science of society that sought "its principal instrument in biology, remaining dismissive or ignorant of economy and linguistics" (Canguilhem 1994, 87). Foucault, in other words, would have benefitted from attending more closely to Comte because the latter provides an "exemplary case of an empirical treatment of the unrelinquished transcendental project" (Canguilhem 1994, 87).

In this chapter, Comte's approach is the paradigmatic example of what I mean by a "foundational" ecological model. Merleau-Ponty's approach, however, is not at all like Comte's. It belongs to a critically different strand of ecological models, which I will characterize as "relational." How, then, do relational models differ from foundational models such as Comte's?

2.4 Relational Ecological Models

Comte introduced the ecological motif—his concern with "the organism and its medium"—as part of his examination of biology in *Course on Positive Philosophy* (Comte 2001). As Canguilhem remarks, the notion of milieu (environment, medium) had been imported from mechanics into biology in the eighteenth century, and Comte's use of this notion remained dominated by its initial mechanical signification (Canguilhem 2008, 99, 101). Furthermore, even though Comte, when considering the human species, was on the brink of formulating a reciprocal conception of the relationship between the organism and the environment, he refused to extend this reciprocity to the living in general, holding the action of the living on the milieu to be negligible (Canguilhem 2008, 102). The hallmark of *relational* ecological models, by contrast, is that they *do* factor in the organism's action on the environment, by conceiving the relationship between organism and environment as genuinely reciprocal.

A seminal contribution to the relational strand of ecological models was made by Jakob von Uexküll, who coined the notion of *Umwelt*. Canguilhem explicates this notion as "the milieu of behavior proper to a certain organism" (Canguilhem 2008, 111), implying that different kinds of organisms have different *Umwelten*, even though they share the same geographical environment. However, throughout its history, the relational version of the ecological motif has evolved and transformed considerably. While von Uexküll's original organism/environment model was static and harmonious, envisioning organisms as perfectly adjusted to their environments and life itself as based on fixed laws (von Uexküll 1926, 84), subsequent thinkers such as Goldstein and Canguilhem critically adjusted the approach by replacing von Uexküll's static model with dialectical models. As a result, the relationship between organism and environment was now seen as an ongoing confrontation. The dialectical model helped explain health and disease: A healthy organism is *more than normal*, it does more than simply adjust itself to the demands of the environment. It has a *normative capacity*, being capable of following new norms of life. The rehabilitation of a sick organism thus corresponds to its capacity to gain a new "individual norm" that guarantees the new order (Canguilhem 1991, 183; Goldstein 1995, 333). Merleau-Ponty, on his side, developed a chiasmatic model, where the body is simultaneously the site of exchange with the world and the "measure of being" (Merleau-Ponty 1968, 215 and 1973, 124).[1]

I will now turn to Simondon, the philosopher who was lucky enough to count both Canguilhem and Merleau-Ponty among his teachers, and whose work can be positioned at the apex of the conceptual trajectory just suggested.

3. Simondon and the Notion of Adaptive Mediator

In their efforts to vindicate the rationality of life and sensibility, Canguilhem and Merleau-Ponty, each in their own way, emphasize that living beings are not machines. A crucial next step in the development of the (relational) ecological model[2] is made by Gilbert Simondon, who extends this point to *technical* objects by contending—perhaps somewhat surprisingly—that *not even machines are machines* (in the usual mechanical sense). This step is crucial, because it relieves the ecological model of its prior dependence on the biological domain, while at the same time acknowledging the rationality of living beings *and* technical beings.

Simondon is primarily known for his notions of individuation and technicity. The theory of individuation was developed in his doctoral thesis (Simondon 2013), which he defended in 1958. His approach to technology was developed in a supplementary thesis, translated to English as *On the Mode of Existence of Technical Objects* (Simondon 2017). In this work, Simondon treats technical objects as "individuals" in the terms laid out in the main doctoral thesis. This implies that he treats technical individuals—or "machines" as he also calls them—as beings that undergo an ontogenetic development much akin to that of living beings. *Mode of Existence* has been much celebrated for its genetic ontology of machines, which elucidates the being of technical objects through a study of their *genesis*—a process of becoming that Simondon refers to as "technical individualization" (Simondon 2017, 63). What has been less remarked upon, is that *Mode of Existence* also provides an ample number of clues to the implications of this ontology for epistemology, especially regarding how the genetic ontology of machines gives a new and unprecedented prominence to technology in knowledge. These clues relate to the roles of technical objects as "mediators" or "intermediaries" (Simondon uses both terms interchangeably).

As I have argued elsewhere (Hoel 2020), a theory of technical mediation can be drawn from these clues. The latter theory complements the theory of technical individualization in that it resonates with the broader scope of Simondon's philosophy of technology, which seeks to elucidate and deepen "the relation which exists between nature, man, and technical reality" (Simondon 2017, xiii). For, as we shall see, the notion of technicity—Simondon's term for technical objects considered in their efficacy and operational functioning—has repercussions far beyond the technical domain, narrowly construed. In Simondon's view, the true philosophical significance of technical objects resides in their power to induce "phase shifts of man to the world" (Simondon 2017, xvii), and hence, to broker the conditions of human life *and* the conditions of knowledge. Thus, in line with Canguilhem and Merleau-Ponty, Simondon insists that, "[i]n reality there exists a great kinship between life and thought" (Simondon 2017, 62). The key to this kinship is the broadened notion of technicity.

The point of this section, then, is to show that the Simondonian machines are promising candidates for what Ian Hacking calls "organizing concepts" (Hacking 1999b, 65).

These organizing concepts, though, are a strange lot, since they operate by creating a highly specific environment around themselves, which they come to depend on for their operation *and* for their further development.

3.1 Simondon's General Theory of Individuation

Simondon's main doctoral thesis opens by commenting on how his theory of individuation differs from established accounts of the living being considered as an individual. The problem with the received accounts, whether of the "substantialist" or the "hylomorphic" variety, is that they neglect the stage of individuation by treating the individual as a given. Simondon, by contrast, sets out "*to understand the individual from the perspective of the process of individuation*" (Simondon 1992, 300, original emphasis). Instead of evoking the traditional notions of substance, matter, and form, Simondon conceives the individual in terms of systems and phases, borrowing the idea of "metastable equilibrium" from modern physics (1992, 301). Established accounts of the individual are inadequate because, lacking the idea of metastability, they recognize nothing but "instability and stability, movement and rest" (1992, 302). Simondon, by contrast, sees the individual as a unit of becoming that undergoes a stepwise evolution that occurs through a series of inventive leaps whereby the individual enters new phases in its development. Crucially, however, the individual is not a separate reality, existing in and by itself. It always forms part of a larger "metastable system" (1992, 302).

It is important to note that "being," for Simondon, fundamentally includes an energetic, vitalist dimension that is missing in received accounts of the individual. To firmly grasp the nature of individuation, being must be considered, not as a substance, matter or form, but as a *system in tension*—a system, that is, which "harbors a certain incompatibility with itself, an incompatibility due at once to forces in tension as well as to the impossibility of interaction between terms of extremely disparate dimensions" (Simondon 1992, 300). Individuation, then, is understood as a "partial and relative resolution" manifested in such a system, and the individual, accordingly, as a "relative reality, occupying only a certain phase of the whole being in question" (1992, 300). Thus conceived, the individual is a precarious entity born out of tensions. Furthermore, for each successive, individuating resolution of the system, a new metastable phase is initiated that releases new potentials for further transformations—which is why Simondon refers to individuation as a "mediate process of amplification" (1992, 304). In this way, the process of individuation attests to a "capacity beings possess of falling out of step with themselves," and "of resolving themselves by the very act of falling out of step" (1992, 300–301).

To better grasp what is at stake in these abstract (and at times peculiar) formulations, it is helpful to remember that, for Simondon, the paradigm example of an individual is a living organism (say, an earthworm or a human being) undergoing development. Simondon considers living being in its process of becoming, seeking to grasp the genesis of the individual in its unfolding. Moreover, in line with the ecological motif, Simondon sees the living organism as intrinsically interwoven with its environment. The organism *forms a joint system with its environment*, meaning that it cannot be properly understood in isolation from its associated milieu (I will return to the notion of associated milieu in the next subsection). It is in this sense that the individual is but a partial resolution of a larger system of being. At the same time, while the organism fully exists in every phase of its development, it remains a relative reality in the sense that the current state of the system never exhausts what the individual can be. The individual has the capacity to change, to negotiate its terms and conditions of existence. The negotiation happens through a process of mediation that initiates a new phase in the system of being, releasing new potentials for action. Thus, while he draws on the notion of metastability from modern physics, Simondon's account of individuation also deeply resonates with Goldstein's and Canguilhem's accounts of the normative capacity of living beings.

3.2 Technical Individualization

While Simondon's theory of individuation is modelled on living beings, it is brought to bear on a much broader range of beings, including technical objects. Even so, Simondon never goes so far as to identify living beings and technical beings. There is a critical difference between the two, which turns on the fact that technical beings owe their origin to human acts of invention. The individuation of technical objects, therefore, is explicated as a process of "concretization" whereby the technical object comes to *approximate* the mode of existence of a living being (Simondon 2017, 25, 29).[3] In Simondon's terms, an "evolved" technical object is more "concrete" than a "primitive" technical object in that its elements and forces are more integrated, approximating the integration of organs in a living body—but also, crucially, in that it has formed a joint system with its surroundings, approximating the vital, reciprocal linkages between a living being and its environment.

There is more to be said about the technical object and how it relates to its environment. A key aspect of Simondon's philosophy of technology is that he approaches the technical object in operational terms as a "being that functions" (Simondon 2017, 151). It is, above all, the operational take on technical objects (including the idea of metastable system) that allows Simondon's approach to break new epistemological ground. For even though the technical object starts out as an "abstract" and "artificial" being, in the course of concretization, it loses some of its artificial character by forming part of a "system of causes and effects that exert themselves in a circular fashion" (2017, 49). The operational take also implies that the technical *individual*—the technical object considered in its individuality or specificity—is not this or that technical object. The technical individual or machine exists, rather, "as a specific type obtained at the end of a convergent series" that "goes from the abstract to the concrete mode" (2017, 28-29). The technical individual or machine is regarded, more precisely, as a certain "schema of functioning," which can be "recognized by the fact that it remains stable across the evolving lineage" (2017, 26, 45, 46). To illustrate what he means, Simondon gives the example of an automobile engine.

The factor deciding whether two engines are the *same* technical individual (in the sense of belonging to the same convergent, evolutionary series of related technical objects), is not the mere fact that both are used to push a car forward. Rather, the decisive factor is whether the two engines operate by the same "regime of causality" (2017, 26).

Another key aspect relates to concretization as a process of adaptation. Also in this case, the operational approach (including metastability) breaks new ground, clearing the way for what Simondon calls "relational adaptation" (Simondon 2017, 57). In Simondon's view, the technical object is seated at the meeting point between two environments that are not completely compatible: the "technical milieu" and the "geographical milieu" (2017, 55). During concretization, the technical object comes to be integrated into both environments at once. In the process, the two worlds start to act upon each other via the technical object, which in this way serves to establish "a reciprocal relation of causality between the technical world and the geographical world" (2017, 56). This is to say that the process of concretization is not one of adapting to a pre-given environment. Adaptation-concretization is considered, rather, as a process that "conditions the birth of a milieu rather than being conditioned by an already given milieu" (2017, 58). What it conditions, more precisely, is the birth of a "third techno-geographic milieu," which "mediates the relation between technical, fabricated elements and natural elements, at the heart of which the technical being functions" (2017, 58, 59). The technical object operates, in other words, by calling forth its own "associated milieu," which in turn is "a condition of possibility of the technical object's functioning" (2017, 58, 59). In Section 4 we will see an example of such an adaptation-concretization process, where a technical object, in our case an MRI machine, calls forth and sustains a highly specific associated milieu, without which there would be no image contrast and hence no images to reveal medically relevant features about the patient body.

What all this amounts to is that the evolution of a Simondonian machine is characterized by a strange self-conditioning: Even though it is invented, and in that sense, artificial, as soon as the machine has formed a joint system with its environment, it takes on a life of its own, developing in ways unforeseen by its inventor(s). This relative autonomy has to do with how the machine "creates its own associated milieu from itself and is really individualized in it" (Simondon 2017, 59), how it calls forth a highly specific environment "that conditions it, just as it is conditioned by it" (2017, 59). It is on the background of this relational notion of adaptation, then, and the corresponding accounts of the relative autonomy and strange self-conditioning of machines, that Simondon characterizes technicity as "both the result and principle of genesis" (2017, 170).

3.3 Technical Mediation

While the theory of technical individualization focuses on the genesis *of* technicity, the theory of technical mediation focuses instead on the genesis that occurs *on the basis of* technicity (Simondon 2017, 171)—on how the technical object participates in the individuation of beings *other* than itself. In both theories, the associated milieu plays a critical role. However, while the former theory emphasizes the stepwise changes that the technical object undergoes in the course of its evolution, the latter theory addresses how the technical object intervenes into the human-world relationship by initiating and sustaining an intermediate structured world, which, according to Simondon, is "a stable mixture of the human and the natural" (2017, 251).

To see more clearly how the theory of technical mediation suggested here amounts to a theory of knowledge, it is useful to remind ourselves of Canguilhem's definition of knowledge as "a general method for the direct or indirect resolution of tensions between man and milieu" (Canguilhem 2008, xviii). It is precisely along these lines that Simondon considers the role of technical mediators: They solve problems by resolving tensions between human and world. Simondon's starting point is that "man and the world form a vital system, comprising the living thing and its milieu" (2017, 168). Crucially, the relationship between human and world is not fixed, but itself subject to development. Technicity, then, is understood to intervene in this development by accomplishing a structural reorganization of the human-world system, which "provisionally resolves the problems posed by the primitive and original phase of man's relationship to the world" (2017, 169). The use of a technical mediator, in other words, brings about a new relative situation of human and world, preparing a new *readiness for action* that was not there (at least not in the same way) in the less evolved human-world system.

Simondon's ideas about mediators are further developed in a 1965–1966 lecture series on imagination and invention (Simondon 2014). The lecture series approaches mediators in broad terms, characterizing them in terms of their "image-value" (2014, 12). The guiding idea of the lecture series is that "everything that intervenes as an intermediary between subject and object can take on the value of an image and play the role of prosthesis, at once adaptive and restrictive" (2014, 12).

The lecture series starts out by considering the images involved in psychological activity. In contrast to the theories of the imagination prevailing at the time (most prominently Sartre 1962 and 1972), Simondon's images are not on the side of the subject, not identified with consciousness, not reducible to human intention, and certainly not opposed to perception. Instead, Simondon refer to them as "motor images": anticipatory behavioral dispositions of the body (or parts of the body) that prepare the living being for its encounter with the environment, and that facilitate a "real coupling" between the two, allowing them to form a joint system (Simondon 2014, 19–20, 92). In this view, living beings come equipped with a reserve of "schemas of conduct" that coordinate and guide their actions in characteristic ways, as exemplified in the instinctual behaviors of animals (2014, 19, 32–33). However, inherited motor images are only "partial programs of behavior" in need of refinement through experience, systematization and innovation. Motor images, in other words, undergo an evolution that takes the form of an "amplification cycle," which serves to install "new anticipations in the long run" (2014, 62). Again, Simondon's approach is in consonance with Canguilhem's idea of living beings striving to gain new norms of life. But this time it also deeply resonates with Merleau-Ponty's analysis of the body schema, including the Merleau-Pontian notions of motor habit and perceptual habit. Thus, much along the lines of Canguilhem and Merleau-Ponty, Simondon vindicates the rationality of living bodies—only now in terms of an amplification-dynamic much akin to technicity—an *originary technicity*, we could say, which is always already at work in living matter.

While motor images are the most elementary of images, they are not the most exemplary. In Simondon's view, the paradigmatic example of adaptive mediation is a situation where a problem is solved through the use of what he calls an "object image"—of some kind of external object (found or fabricated) that serves as an "adaptive mediator" (Simondon 2014, 141, 142). An example of such an image object would be a winch used to move a heavy load. By using a winch, a human being is able to handle the load *as if* she were much stronger than she actually is. Thus, as in the case of motor images, the adaptive mediator *amplifies* the human-world system by *realizing a transfer to a new level*—only this time, the inventive leap is more considerable. Furthermore, object images differ from motor images in that they exist independently as detached cultural artifacts. Due to their detached existence, object images, such as tools, instruments and machines, can be used by other humans far from the time and place of their creation. In Simondon's view, this means that object images realize the transfer function more perfectly than motor images, in the sense of having stronger cumulative and collective world-building effects.

In the following section, I will consider MRI as an object image in Simondon's sense, and thus, as an *adaptive mediator*.

4. MRI as an Adaptive Mediator

Science studies have granted much attention to the role of imaging and visualization in knowledge formation. As indicated by the titles of landmark publications such as *Representation in Scientific Practice* (Lynch and Woolgar, 1990) and its follow-up *Representation in Scientific Practice Revisited* (Coopmans et al., 2014), the prevailing tendency is to approach scientific and medical images as representations.[4] This implies, first, that the notion of image tends to be identified with the *result* of the imaging process, the characteristic gray-level (and sometimes chromatic) visual displays shown on computer screens, light boxes and similar; and second, that the majority of studies focus on the image-observer relation and the practices involved in analyzing and interpreting such visual displays (e.g., Alač 2011)—on the assumption that the problematic arises in this stage of the process. This assumption is clearly warranted, since there are real issues involved in the interpretation of scans. Even for highly trained observers, the presence of a specific object is not always obvious. Besides, different observers may use different criteria to establish what is seen, sometimes resulting in disputes over conflicting interpretations (e.g., Rosenberger 2011).

For all that, a new line of research is emerging that takes a broader approach by factoring in machine agency (e.g., Vertesi 2015). Contributing to the second line of research, the approach developed here seeks to supplement the established studies by calling attention to the preceding stage of the process, that of image *acquisition*. This chapter, in other words, is guided by the assumption that the problematic arises already—and in some respects more decisively—in this preceding step, where the machine installs and stabilizes a highly specific associated milieu that grounds object visibility *and* viewing conditions in non-trivial ways.

4.1 MRI and the Generation of Image Contrast

The task of medical imaging systems is to reveal medically relevant features of the human body by translating specific tissue characteristics into different shades of gray or color in the image (Sprawls 1995, 1, 3). MRI is a variety of tomographic imaging, which produces images of selected planes of tissue in the patient body. In clinical examinations, MRI is often the preferred imaging modality, since it provides excellent soft tissue discrimination (Westbrook and Talbot 2019, 24).

MRI technology relies on the magnetic behavior of hydrogen nuclei or protons, which exist in great quantities in living bodies, especially water and fat. Since the magnetic behavior of such protons vary systematically depending on the tissue, MRI is used to map the boundaries between different tissue types, and also, crucially, between healthy and pathological tissues. During the MRI scanning process, differences between tissues are indicated by differences in signal intensity, which show up on MRI scans as differences in brightness (gray level)—areas of high signal appearing bright in the image, and areas of low signal appearing dark in the image (Westbrook and Talbot 2019, 31). These differences in signal intensity and/or brightness are referred to as “image contrast” (McRobbie et al. 2003, 30). In the MRI literature, the image contrast resolution is considered *adequate* to the extent that it ensures optimal differentiation of tissue structures and reveals pathology. This means that the visibility of a certain object (say, a tumor) depends on whether it has sufficient contrast relative to surrounding tissues.[5] For the purposes of this chapter, it is important to note that the degree of contrast in the image depends on the characteristics of the tissues examined *and* the characteristics of the imaging system (Sprawls 1995, 3). Restated in the terms of Simondon: The image contrast resolution depends on the characteristics of natural *and* technical elements and forces, as these come to be stabilized in a recurrent regime of reciprocal causalities.

4.2 The MRI Machine and Its Associated Milieu

The Simondonian idea that there are technicities on the side of nature, implies that living matter is more than an aggregate of simple qualities. Endowed with technicities, natural elements can be thought of as specific “capacities for producing or undergoing an effect in a determinate manner” (Simondon 2017, 75). Conceived along these lines, hydrogen protons can be seen as micro-scale motor images that express a certain behavioral potential. Before the intervention of MRI technology, the hydrogen protons in the body are randomly spinning along their axes. However, as soon as the patient is positioned in the scanner, the protons line up with the strong magnetic field in the bore of the scanner. To generate the signal, the scanner produces a rapidly repeating sequence of radiofrequency pulses applied at 90 degrees to the main magnetic field. These pulses "excite" the protons, forcing them to absorb energy and spin in a different direction. Each time the radiofrequency pulses are turned off, the protons start to "relax," releasing their excess energy as they realign with the main magnetic field (Westbrook and Talbot 2019, 14). In releasing this energy, the protons give off an electromagnetic signal that is detected by the scanner. This signal is then digitized as a function of time and translated into an image matrix (McRobbie et al. 2003, 47, 57, 58).

This chapter is not the place to dwell on the intricacies of MRI physics. Still, before we leave the world of excited and relaxing protons behind, a few more observations need to be made about the relaxation process, since these have direct bearing on my argument. Depending on the tissue composition and the strength of the magnetic field, different types of tissue have different relaxation times, and the same goes for healthy and pathological tissues. There are two ways of measuring the time it takes for the hydrogen protons in a certain tissue to relax. The first, which in the MRI literature is referred to as "T1 recovery time," measures the time it takes for protons to recover their magnetization in the longitudinal direction; and the second, "T2 decay time," measures the time it takes for the spins to lose their coherent magnetization in the transverse direction (Westbrook and Talbot 2019, 26–28). Thus, to be more precise, it is the tissue-specific time constants associated with T1 and T2 relaxation that form the basis of image contrast in MRI.

It should be clear even from this rough sketch of MR image acquisition that the MRI machine quite literally conditions the existence of a highly specific associated milieu that it depends on for its functioning. But the critical point here is that it *also depends on this associated milieu for its capacity to reveal something about the body undergoing examination*. Thus, while in the MRI literature the T1 and T2 relaxation times are often presented as "inherent to the body's tissues" (Westbrook and Talbot 2019, 25), it transpires from the earlier discussion that the differential behaviors that the hydrogen protons exhibit as they relax at different rates in different tissues, are not found in the natural state of the body. It is not until the magnetic moments of the protons have been appropriately modified—*concretized*, we could say—first, by coming under the influence of the strong external magnetic field of the scanner; and second, by being subjected to systematic manipulation by radiofrequency pulses, that the behaviors of the protons become sufficiently stable and law-like to be used as reliable measures in the generation of image contrast.

4.3 Differential Principles of Individuation

As we have seen, MR image contrast is based on the T1 and T2 relaxation times as these vary systematically between types of tissues. This baseline resolution can be further modified through the tweaking of the scan parameters. This tweaking relates to the imaging variables selected by the operator, including the choice of pulse sequences and the timings of these sequences. In neuroradiology, where MRI is used in the diagnosis of tumors and other lesions of the brain, the choice of repetition time, echo time, and other factors affects which features of the patient's brain are singled out. Pulse sequences with short repetition time and short echo time, for example, tend to enhance the T1 differences between tissues, resulting in "T1-weighted images" in which fat-based tissues appear bright in the image; whereas pulse sequences with long repetition time and long echo time tend to enhance T2 differences, resulting in "T2-weighted images" in which fluids appear bright (McRobbie et al. 2003, 32, 33). Since pathologies of the brain are frequently associated with accumulation of fluids, a lesion such as a tumor will typically be more visible in a T2-weighted image than in a T1-weighted image. In addition to T1-weighted and T2-weighted sequences, routine clinical examinations also typically include other sequences that allow for even more subtle optimizations of image contrast. Furthermore, in addition to the tweaking of scan parameters, the contrast resolution of MRI can also be modified by the use of various contrast agents that are injected into the patient's bloodstream.

The aim of the modifications mentioned is to increase the adequacy of the contrast resolution, in order to enhance the ability of the imaging system to single out precisely those features that are clinically relevant—say, whether a certain tissue is tumorous or not, and if it is, the size and location of the tumor. At no point, however, will the distinctions made by the MRI machine fully coincide with those intended by the clinician. The MRI machine, of course, knows nothing about tumors, and when in operation, it relentlessly performs in accordance with its schema regardless of human concerns about health and disease. Still, there is rationality and knowledge involved, relating to how the MRI machine enacts an individuating resolution of tensions manifested in its highly specific associated milieu, which can be characterized operationally in terms of a recurrent regime of reciprocal causalities that critically includes the dynamics of excitation and relaxation just accounted for. Thus, when considered from the point of view of technical mediation, the operational schema of the MRI machine takes on the role as a principle of genesis (or individuation) of *other* beings—in our case, as a principle for distinguishing between different tissue types. Moreover, the example of MRI as employed in the diagnosis of brain tumors, allows us to further specify this schema as a *differential* principle of individuation, since the MRI scanner, as we have seen, operates by enacting a divergence between foreground and background (hence the term "contrast resolution"). In fact, on closer scrutiny, the operational schema of MRI gives rise to a whole *family* of related differential principles—since for each setting of the system parameters there is a new principle of individuation, a new distribution of foreground and background, and ultimately, a new method for individuating (differentiating and articulating) the object of knowledge.

In what sense, then, is the MRI machine an "image" in the Simondonian outlook? It is an image, first, in that it is an adaptive mediator that realizes a transfer to a new level of the human-world system, instituting a highly specific milieu of individuation that concretizes the phenomena of interest and provokes them to exhibit law-like behaviors that translate into systematic patterns that are apprehensible to human senses and that can be exploited for medical purposes. The MRI machine is an image, second, in that it has a *figural* dimension, orchestrating targeted differential orderings of the world where certain features are delineated and made to stand out as *figures* (appearing bright in the resulting gray-level image matrix), while other features blend into the background or do not show at all.

5. Styles of Objectivity

The point of examining the workings of MRI in this much detail is to show that the MRI machine performs concept work of sorts. Put another way, MRI is chosen here to elucidate the sense in which Simondonian machines are promising candidates for organizing concepts. Furthermore, what MRI helps demonstrate is that technicity, in its role as adaptive mediator, acts as a "force of divergence" (Simondon 2017, 171): The machine operates by enacting specific *figure-ground resolutions* of its corresponding metastable system, which in turn serves to individuate (concretize and further articulate) the phenomena under scrutiny. This implies that the machine *intervenes* into the phenomena it examines, by acting as a *differential* organizing principle that guides the individuation of the object of knowledge in one direction rather than another. The machine-cum-mediator, therefore, can be characterized in terms of its *distinct style of individuating and revealing the object*.

By conceiving machines in this way, Simondon extends to the machine an insight that Merleau-Ponty had already made for the perceiving body, namely, that "perception already stylizes" (Merleau-Ponty 1973, 60). Merleau-Ponty's argument goes like this:

> Style exists (and hence signification) as soon as there are figures and backgrounds, a norm and a deviation, a top and a bottom, that is, as soon as certain elements of the world assume the value of dimensions to which subsequently all the rest relate and through which we can point them out.
>
> (Merleau-Ponty 1973, 61)

What, then, becomes of the object of knowledge when its boundaries qua object are seen as differentially enacted? In his discussion of motor images, Simondon gives us a clue. The existence of motor images, he maintains, allows us to analyze the object with more precision by conceiving it as a "mode of relation between the organism and the environment" (Simondon 2014, 29). This idea, that there are "modalities of the object" (Simondon 2014, 33) that somehow correspond to the anticipatory behavioral disposition of the organism, is pivotal to relational ecological models. It was already assumed in the notion of *Umwelt*—as when Jakob von Uexküll remarks that the "life-path of an earthworm" is composed of nothing but "earthworm things" (von Uexküll 1926, 307). For all that, the ecological motif has transformed significantly from von Uexküll's fixed life-tunnels to Simondon's open machines conditioning and being conditioned by their associated milieus. But more than the dynamic idea of co-conditioning, the difference that makes a difference is Simondon's idea of an *operational coupling* between the organism/machine and the environment.

This is also the point where the *ecological* relationality proposed here differs most markedly from the *poststructuralist* relationality that informs ANT.

5.1 Poststructuralist Relationality

On the face of it, there are strong affinities between the ecological and poststructuralist notions of relationality. Both notions emphasize heterogeneity, materiality and process, both take issue with foundational divisions and both deal with enactments and the precarious generation of realities. Yet they differ on a crucial point—the point of "how it is that everything hangs together," to borrow John Law's turn of phrase in his much-quoted chapter on ANT and material semiotics (Law 2009, 145). In the chapter mentioned, Law defines ANT as "a disparate family of material-semiotic tools, sensibilities, and methods of analysis that treat everything in the social and natural worlds as a continuously generated effect of the webs of relations within which they are located" (Law 2009, 141). He goes on to characterize ANT as "an empirical version of poststructuralism" and actor networks as "scaled-down versions of Michel Foucault's discourses or epistemes" (Law 2009, 145). As Law makes clear, ANT is committed to a material-semiotic or poststructuralist notion of relationality that erodes "ontological distinctions" and levels "divisions usually taken to be foundational" (Law 2009, 147).

The commitment to poststructuralist relationality has proven to be analytically productive. It gives ANT a critical edge, allowing it to "follow surprising actors to equally surprising places" (Law 2009, 147). However, as Law admits, this commitment is also the source of much debate: "as with Foucault, there is a powerful if controversial nonhumanist relational and semiotic logic at work" (Law 2009, 147). Apart from characterizing it as "material-semiotic" and "poststructuralist," Law does not delve further into what kind of relational and semiotic logic we are dealing with here. But what he is alluding to is a certain model of relationality that harkens back to structuralist linguistics, and that took its impetus from Ferdinand de Saussure's doctrine of the arbitrary character of linguistic signs (de Saussure 1959). Clearly, this model has changed considerably since the heyday of structuralism. In today's material-semiotic approaches, the static Saussurean systems of differences have been replaced by multifarious dynamic networks, relations now being considered as materially and discursively heterogeneous. Nonetheless, there is something distinctly *Saussurean* about how ANT conceives of networks and relations, as when it assumes that "nothing has reality or form outside the enactment of those relations" (Law 2009, 141). This, then, is why ANT's commitment to poststructuralist relationality is a source of controversy: By warranting the treatment of everything as *nothing but contingent relational effects*, it evokes the specter of relativism.

The replacement of necessary causes by contingent effects also involves other challenges. While the commitment to poststructuralist relationality allows ANT to explore a "non-foundational world" where "nothing is sacred and nothing is necessarily fixed" (Law 2009, 148), it raises the issue of what to focus on to effectively study the material practices that generate the social—given that all criteria that help distinguish the more relevant from the less relevant have now been obliterated. As noted by Law, a common objection leveled against ANT is that it gets lost in material minutiae and fails to attend sufficiently to what is important (Law 2009, 148). It also raises the issue of how to account for the durability or stability of the networks studied. In the words of Law: "what might replace the foundations that have been so cheerfully undone? Is it possible to say anything about network-stabilizing regularities, or are we simply left with describing cases, case by case?" (Law 2009, 148).

5.2 Ecological Relationality

The ecological model differs from the poststructuralist model in how it conceives of systems and relations. More than a mere network of relations, Gilbert Simondon's paradigmatic system is a *metastable system* in the sense outlined previously: It is a "supersaturated" milieu full of potential—a "being" that is more than substance, matter or form in that it "exists at a higher level than the unit itself, which is not sufficient unto itself and cannot be adequately conceptualized according to the principle of the excluded middle" (Simondon 1992, 301). This implies that there is always *more* to the system than the elements and forces that, at any given point of time, are actualized in the prevailing regime of tensions. This implies, in turn, that the *reality* of whatever comes to be individuated (concretized and further articulated) in a certain system is never exhausted by the current individuating resolution of this system. There is always *more* to be revealed—say, about the tissues scrutinized by MRI. Furthermore, in the ecological model, relations are *grounded*: They grow out of tensions—as in the case of MRI, out of technical and natural elements and forces in their mutual reaction. This also means that relations are *figural*: They come to be expressed in characteristic figure-ground resolutions—again as in MRI, where the figural resolution of various tissues "relaxing" at different rates translates into a meaningful visual contrast in the resulting image matrix.[6]

The implication of all this is that, even though the milieu of individuation has been conditioned into being by the intervention of some technical artifact (in our case an MRI scanner), there is something *necessary* and *law-like* about the contrast patterns that emerge from this mixed milieu. What we are dealing with here, however, is a strange new breed of *contingent necessities* that are made possible because relations now take the form of operational couplings. For, as Simondon insists, even though the associated milieu is, in a sense, created by the machine, it is not entirely "fabricated" (Simondon 2017, 59). The associated milieu is not entirely fabricated because it "incorporates a part of the natural world that intervenes as a condition of functioning, and is thus part of the system of causes and effects" (2017, 49).

The operational-ecological model also differs from the poststructuralist model in the centrality it accords to apparatuses (broadly conceived). While ANT takes technology seriously by including technical artifacts among the actors that must be accounted for when describing a certain web of relations, there are no adaptive mediators in Simondon's sense. The operational-ecological approach differs in that it treats apparatuses as *conditions* of sorts. This implies that the machine is not just a factor among other factors in the system; it is the factor that *conditions the system into being*, that *sustains it* and *gives it a direction*. Nevertheless, and even though the machine-cum-mediator carries out its task in accordance with its own distinct style of individuation, it does not determine or prescribe in advance the patterns that emerge from its associated milieu. The strange self-conditioning that characterizes the individualization of machines also does the trick when it comes to technical mediation—allowing the machine/mediator to *intervene* into and *transform* phenomena while at the same time *revealing* something about them.

6. Unlocking Epistemologies beyond the Positions of the Science Wars

The bitter controversy over social construction seems to be driven by two great fears that are equally justified: the fear of relativism and the fear of foundationalism. Having consulted Michel Foucault's *The Order of Things*, it is striking to notice the extent to which the opposing camps of the science wars map onto the two kinds of truth identified in Foucault's work: the positivist truth that is *of the same order as the object* and the eschatological truth that is *of the order of discourse*.

The operational-ecological model proposed in this chapter opens a third possibility. When Maurice Merleau-Ponty, in *The Visible and the Invisible*, talks about the need to "situate ourselves within the being we are dealing with," and to put being "back into the fabric of our life,"[7] he is not calling for a return to "actual experience" (as Foucault thinks phenomenology is all about, to judge from his discussion in *The Order of Things*); nor is he calling for a return to "a philosophy of experience, of sense and of subject" (as Foucault suggests by positioning Merleau-Ponty on the "subjectivist" side of the central cleavage in postwar French philosophy). What Foucault seems oblivious about, is that Merleau-Ponty has *already shifted to an operational-ecological conceptual register*, which opens the way for new epistemologies beyond the options outlined in *The Order of Things*. What Merleau-Ponty calls for, then, is a return to the "milieu" in the multiple senses of this word in French: to the *middle*, to the *environment*, to the *medium*. What he asks us to do is to investigate the body not so much as a thing among things as "the measurant of the things" (Merleau-Ponty 1968, 152).

Simondon, likewise, emphasizes the middle: Technicity resolves incompatibilities in the human-world system by instituting a "*middle* order of magnitude" (Simondon 1992, 304, original emphasis) that cuts across existing orders and, in so doing, overcomes the initial absence of communication between the disparate parts of the system. The associated milieu is, per definition, a *third* order that is neither of the *same* order as the object nor of an altogether *different* order. In its role as adaptive mediator, the technical object enables a "convertibility of the human into the natural and of the natural into the human," which leads to "a new relative situation of man and nature" (2017, 251). Through the intervention of the technical object, the "relation of man to nature" takes on "a status of stability, of consistency, making it a reality that has laws and an ordered permanence" (2017, 251).

We have now arrived at the point where Simondon's approach differs most decisively from Gaston Bachelard's. While Bachelard, in his attempt to secure the objectivity of scientific knowledge, felt compelled to accentuate the *artificiality* of scientific phenomenology, in sharp contradistinction to the alleged *naturalness* of primary experience, Simondon suggests instead (as indicated by his account of motor images) that primary experience is *always already amplified*—and hence, not that primary after all. There are no "ready-made" objects anywhere, not even in perception. Moreover, since the convertibility of the human and the natural goes both ways, a similar argument can be made for the objects of modern science: While in most cases they are definitely *realized*—produced or generated—by some intervening instrument or machine, they may not be all that artificial, in spite of indications to the contrary. As suggested by the Simondonian idea of concretization: Even though the technical object starts out as artificial and disconnected, it loses some of its artificial character as soon as it is put to use in the world. By forging operational linkages to a more-than-technical environment, the technical object becomes *real* in a new operational sense of the term: *efficacious*.

It is, above all, the understanding of relations as *operational couplings* that breaks new ground, epistemologically speaking. This observation finds some support in the concluding pages of *On the Mode of Existence of Technical Objects*, where Simondon comes close to launching the technical operation as an alternative paradigm of truth. He starts out by establishing that, since it builds an intermediate reality that leads to a new relative situation of human and world, the technical operation is not "pure empiricism" (Simondon 2017, 251). He proceeds to criticize pragmatist and nominalist approaches for conflating the practical and the operational, and hence, for ignoring that "the technical operation is not arbitrary" (Simondon 2017, 260).

I follow up on this by proposing an operational-ecological model that treats machines as organizing concepts—or what amounts to the same: *apparatuses as material concepts*. The idea of material concepts requires a significant broadening of the notion of rationality as we have come to know it. Certainly, I am not alone in calling for such a broadening; it is, I believe, an emerging trend. Ian Hacking, for example, touches upon the need for a broader approach when he, reflecting upon his choice of terms, comes to realize that the word "reasoning" has "too much to do with mind and mouth and keyboard; it does not, I regret, sufficiently invoke the manipulative hand and the attentive eye" (Hacking 1992, 4). Foucault, on his side, went much further, coming very close to making an operational turn when, in his later work, he attempted to push rationality beyond discourse by recentering his work on the notion of *dispositif* (Foucault 1980, 197). But what does it entail to treat apparatuses as material concepts? Clearly, it would radically change what concepts are—but also, *where* they are and what they do.

Notes

1. For a detailed account of Merleau-Ponty's highly original idea of the body as a standard of measurement, see Hoel and Carusi (2018).
2. In the remainder of this chapter, when I talk about the "ecological model," I mean the *relational* ecological model (if not otherwise indicated).
3. Living beings are for Simondon the paradigm case of an "entirely concrete existence"—a kind of existence that technical objects tend towards but can never fully obtain (Simondon 2017, 51).
4. That said, both publications, and the latter especially, draw on a broad range of different theoretical perspectives.
5. Put more precisely: The ability to see a tumor in the image depends on whether the tumorous tissue gives off a signal that is sufficiently different in intensity relative to the signals given off by the surrounding tissues for it to show up in the resulting images as a visible gray-level difference.
6. It is important to note, here, that "figural" in this context does not necessarily mean "visual." Figure-ground resolutions can also be expressed in other modalities, say, by differences in number values. The resolution is figural, rather, by virtue of being a *differential pattern expressed through contrast*.
7. The whole quote goes like this: "Before the essence as before the fact, all we must do is situate ourselves within the being we are dealing with, instead of looking at it from the outside—or, *what amounts to the same thing*, what we have to do is put it back into the fabric of our life, attend from within to the dehiscence (analogous to that of my own body) which opens it to itself and opens us upon it, and which, in the case of the essence, is the dehiscence of the speaking and the thinking." (Merleau-Ponty 1968, 117–118)

References

Achterhuis, Hans. 2001. *American Philosophy of Technology: The Empirical Turn*. Translated by Robert P. Crease. Bloomington and Indianapolis: Indiana University Press.

Alač, Morana. 2011. *Handling Digital Brains: A Laboratory Study of Multimodal Semiotic Interaction in the Age of Computers*. Cambridge, MA: The MIT Press.

Bachelard, Gaston. 1984 [1934]. *The New Scientific Spirit*. Translated by Arthur Goldhammer. Foreword by Patrick A. Heelan. Boston: Beacon Press.

Bachelard, Gaston. 2002 [1938]. *The Formation of the Scientific Mind: A Contribution to a Psychoanalysis of Objective Knowledge*. Introduced, translated, and annotated by Mary McAllester Jones. Manchester: Clinamen Press.

Canguilhem, Georges. 1991 [1943]. *The Normal and the Pathological*. Translated by Carolyn R. Fawcett in collaboration with Robert S. Cohen. With an introduction by Michel Foucault. New York: Zone Books.

Canguilhem, Georges. 1994 [1967]. "The Death of Man, or Exhaustion of the Cogito?" Translated by Catherine Porter. In *The Cambridge Companion to Foucault*, edited by Gary Gutting, 71–91. Cambridge: Cambridge University Press.

Canguilhem, Georges. 2008 [1965]. *Knowledge of Life*, edited by Paola Marrati and Todd Meyers. Translated by Stefanos Geroulanos and Daniela Ginsburg. New York: Fordham University Press.

Comte, Auguste. 2001 [1830–1842]. *The Positive Philosophy of Auguste Comte*, 2 vols. Bristol: Thoemmes Press.

Coopmans, Catelijne, Janet Vertesi, Michael Lynch, and Steve Woolgar. 2014. *Representation in Scientific Practice Revisited*. Cambridge, MA: MIT Press.

Farocki, Harun. 2004. "Phantom Images." *Public* 29: 12–24.

Fleck, Ludwik. 1979 [1935]. *Genesis and Development of a Scientific Fact*, edited by Thaddeus J. Trenn and Robert K. Merton. Translated by Fred Bradley and Thaddeus J. Trenn. Foreword by Thomas S. Kuhn. Chicago and London: The University of Chicago Press.

Foucault, Michel. 1970 [1966]. *The Order of Things: An Archaeology of the Human Sciences*. Translated from the French. London and New York: Routledge.

Foucault, Michel. 1980. *Power/Knowledge: Selected Interviews and Other Writings, 1972–1977*, edited by Colin Gordon. Translated by Colin Gordon, Leo Marshall, John Mepham, and Kate Sper. New York: Pantheon Books.

Foucault, Michel. 1991 [1978]. "Introduction." In George Canguilhem, *The Normal and the Pathological*, 7–24. New York: Zone Books.

Goldstein, Kurt. 1995 [1934]. *The Organism: A Holistic Approach to Biology Derived from Pathological Data in Man*, with a foreword by Oliver Sacks. New York: Zone Books.

Gutting, Gary. 1989. *Michel Foucault's Archaeology of Scientific Reason: Science and the History of Reason*. Cambridge: Cambridge University Press.

Hacking, Ian. 1982. "Language, Truth and Reason." In *Rationality and Relativism*, edited by Martin Hollis and Steven Lukes, 48–66. Oxford, Blackwell.

Hacking, Ian. 1992. " 'Style' for Historians and Philosophers." *Studies in History and Philosophy of Science* 23, no. 1: 1–20.

Hacking, Ian. 1999a. *The Social Construction of What?* Cambridge, MA and London: Harvard University Press.

Hacking, Ian. 1999b. "Historical Meta-Epistemology." In *Wahrheit und Geschichte*, edited by Wolfgang Carl and Lorraine Daston, 53–77. Göttingen: Vandenhoeck & Ruprecht.

Hacking, Ian. 2012. " 'Language, Truth and Reason' 30 Years Later." *Studies in History and Philosophy of Science* 43, no. 4: 599–609.

Hoel, A. S. Aurora. 2020. "Images as Active Powers for Reality: A Simondonian Approach to Medical Imaging." In *Dynamis of Images: Moving Images in a Global World*, edited by Emmanuel Alloa and Chiara Cappelletto, 287–310. Berlin: De Gruyter.

Hoel, A. S. [Aurora] and Annamaria Carusi. 2018. "Merleau-Ponty and the Measuring Body." *Theory, Culture & Society* 35, no. 1: 45–70.

Ihde, Don. 2003. "If Phenomenology Is an Albatross, Is Post-phenomenology Possible?" In *Chasing Technoscience*, edited by Don Ihde and Evan Selinger, 131–144. Bloomington: Indiana University Press.

Ihde, Don. 2009. *Postphenomenology and Technoscience: The Peking Lectures*. Albany: State University of New York Press.

Ihde, Don. 2015. "Preface: Positioning Postphenomenology." In *Postphenomenological Investigation: Essays on Human-Technology Relations*, edited by Robert Rosenberger and Peter-Paul Verbeek, vii–xvi. Lanham, Boulder, New York, and London: Lexington Books.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Construction of Scientific Facts*. Beverly Hills: Sage Publications.

Law, John. 2009. "Actor Network Theory and Material Semiotics." In *The New Blackwell Companion to Social Theory*, edited by Bryan S. Tumer, 141–158. Chichester and Malden, MA: Wiley-Blackwell.

Lynch, Michael, and Steve Woolgar (eds.). 1990. *Representation in Scientific Practice*. Cambridge, MA: The MIT Press.

McRobbie, Donald W., Elizabeth A. Moore, Martin J. Graves, and Martin R. Prince. 2003. *MRI: From Picture to Proton*, second edition. Cambridge, UK: Cambridge University Press.

Merleau-Ponty, Maurice. 1963 [1942]. *The Structure of Behavior*. Translated by Alden L. Fisher. Boston: Beacon Press.

Merleau-Ponty, Maurice. 1968 [1964]. *The Visible and the Invisible*, edited by Claude Lefort. Translated by Alphonso Lingis. Evanston: Northwestern University Press.

Merleau-Ponty, Maurice. 1973 [1969]. *The Prose of the World*, edited by Claude Lefort. Translated by John O'Neill. Evanston: Northwestern University Press.

Mitchell, W. J. T., and Mark B. N. Hansen. 2010. *Critical Terms for Media Studies*. Chicago: University of Chicago Press.

Peters, John Durham. 2015. T*he Marvelous Clouds: Toward a Philosophy of Elemental Media*. Chicago: University of Chicago Press.

Pickering, Andrew. 1984. *Constructing Quarks: A Sociological History of Particle Physics*. Chicago: Chicago University Press.

Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.

Rheinberger, Hans-Jörg. 2005. "Gaston Bachelard and the Notion of 'Phenomenotechnique.'" *Perspectives on Science* 13, no. 3: 313–328.

Rosenberger, Robert. 2011. "A Case Study in the Applied Philosophy of Imaging: The Synaptic Vesicle Debate." *Science, Technology & Human Values* 36, no. 1: 6–32.

Sartre, Jean-Paul. 1962 [1936]. *Imagination: A Psychological Critique*. Translated by Kenneth Williford and David Rudrauf. London and New York: Routledge.

Sartre, Jean-Paul. 1972 [1940]. *The Psychology of the Imagination*. London: Routledge.

Saussure, Ferdinand de. 1959 [1916]. *Course in General Linguistics*, edited by Charles Bally and Albert Sechehaye, in collaboration with Albert Reidlinger. Translated by Wade Baskin. New York: Philosophical Library.

Sciortino, Luca. 2017. "On Ian Hacking's Notion of Style of Reasoning." *Erkenntnis* 82: 243–264.

Simondon, Gilbert. 1992. "The Genesis of the Individual." In *Incorporations*, edited by Jonathan Crary and Sanford Kwinter. Translated by Mark Cohen and Sanford Kwinter, 297–319. New York: Zone Books.

Simondon, Gilbert. 2013. *L'individuation à la lumière des notions de forme et d'information*, Grenoble: Millon,

Simondon, Gilbert. 2014. *Imagination et invention*, 1965–1966. Paris: Presses Universitaires de France.

Simondon, Gilbert. 2017 [1958]. *On the Mode of Existence of Technical Objects*. Translated by Cecile Malaspina and John Rogove. Minneapolis and London: University of Minnesota Press.

Sprawls, Perry Jr. 1995. *Physical Principles of Medical Imaging*, 2nd edition. Madison, Wisconsin: Medical Physics Publishing.

Uexküll, Jakob von. 1926. *Theoretical Biology*. New York: Harcourt, Brace & Company.

Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. Translated by Robert P. Crease. University Park: The Pennsylvania State University Press.

Vertesi, Janet. 2015. *Seeing Like a Rover: How Robots, Teams, and Images Craft Knowledge of Mars*. Chicago: University of Chicago Press.

Westbrook, Catherine, and John Talbot. 2019. *MRI in Practice*, fifth edition, Oxford: Wiley-Blackwell.

CHAPTER 7

ENGINEERING KNOWLEDGE

WYBO HOUKES AND ANTHONIE MEIJERS

1. Introduction

This chapter is concerned with engineering as an epistemic activity, that is, as producing and using knowledge. It deals with, to quote the title of Walter Vincenti's monograph on this topic, *what engineers know and how they know it*. We outline the main existing perspectives on this issue, developed by engineers and by scholars reflecting on their practices, and then offer ingredients for an alternative analysis that combines elements of existing perspectives while avoiding some of their shortcomings.

By way of introducing the main existing perspectives: "Born to Engineer," a campaign launched by the ERA Foundation in 2010, states that "engineers turn ideas into reality" and that "engineers are creative problem solvers" (Born to Engineer 2019). Similarly, according to the Royal Academy of Engineering, engineering "brings ideas to life and turns dreams into reality" and design "turns creativity into real-life solutions, producing products and services" (Royal Academy of Engineering 2018).

At first glance, such statements suggest that generating knowledge is at best a secondary aim of engineering and design, instrumental to its primary aim of shaping reality or solving problems. Even this might be an overstatement. Claims such as "Engineers use maths, science—especially physics— . . . to turn ideas into reality" (Tomorrow's Engineers 2019) suggest that engineers are professional knowledge-*consumers*, albeit perhaps creative ones.

Many disagree. Some claim that engineers and designers are professional knowledge-*producers*, but that their epistemic products are of a special type. For instance, developing statements in a 1979 report by the Royal College of Art, Nigel Cross (1982) argued that there are "designerly ways of knowing" embodied in products and processes. Vincenti starts his book by stating that "technology appears . . . as an autonomous body of knowledge, identifiably different from the scientific knowledge with which it interacts" (Vincenti 1990, 1–2). Others agree in regarding engineering as an epistemic activity but deny that it is identifiably different from science. They might for instance denote both scientific and engineering practices as "technoscience" (e.g., Latour 1987), or argue that all contemporary research involves "Mode-2" knowledge (Gibbons et al. 1994) or "Triple-Helix" collaborations (Etzkowicz and Leydesdorff 2000).

This chapter seeks to advance our understanding of engineering knowledge. In Section 3, we distinguish existing views of engineering knowledge as *subordinating* (3.1), *contrasting* (3.2), or *assimilating* (3.3) it to (natural-)scientific knowledge. After identifying shortcomings and useful elements of each view, the chapter offers ingredients for an alternative analysis. In Section 4, we sketch how the design of high-tech systems involves sets of *epistemic activities* (4.1), resulting in a variety of *rules* (4.2), which (i.e., activities and rules) are governed by a distinctive set of epistemic and *non-epistemic values* (4.3). Throughout, we primarily use one case to illustrate our points: the development of the nuclear-fusion test reactor ITER. In Section 2, we give some background information on this case to provide context for the illustrative details supplied in later sections. Section 5 provides conclusions and some points for further research.

2. Fusion Engineering in ITER

Nuclear fusion releases large amounts of energy if it involves combination of light atomic nuclei. It has been discovered to power stars, and has been used in thermonuclear weapons; both theories of stellar nucleosynthesis and "proof-of-concept" nuclear devices were developed in the 1950s. Around the same time, work started on using fusion for generating electricity. Decades later, this might result in the first functional, net-yield fusion reactor starting operation in 2025. This reactor—ITER—is described as an "experimental tool," "crucial to advancing fusion science" but also as "designed to prove the feasibility of fusion as a . . . source of energy" and "to test the integrated technologies, materials, and physics regimes necessary for the commercial production of fusion-based electricity" (ITER Organization 2019). It is astoundingly complex—as a research project that involves a thirty-five-year collaboration of thousands of people from thirty-five countries and its own monetary unit, and as a technological system.

ITER's central device, a tokamak (see Figure 7.1), is a toroid vacuum chamber that contains electrically charged hydrogen gas—a plasma. This tokamak has an estimated ten million individual parts, many of which need to operate under extreme conditions, such as temperatures of a hundred million degrees Celsius. Heat produced through fusion reactions in the plasma is transported through the wall of the chamber or "vessel." The plasma itself is shaped by magnetic confinement so that it does not touch and damage the walls. For this, the largest superconducting magnetic system ever will be used, containing over 100,000 kilometers of niobium-tin strands, which alone required a six fold increase of global production capacity of this material.

The tokamak configuration is widely accepted as the most promising for producing fusion power. There are, however, alternative configurations—such as the "stellerator"

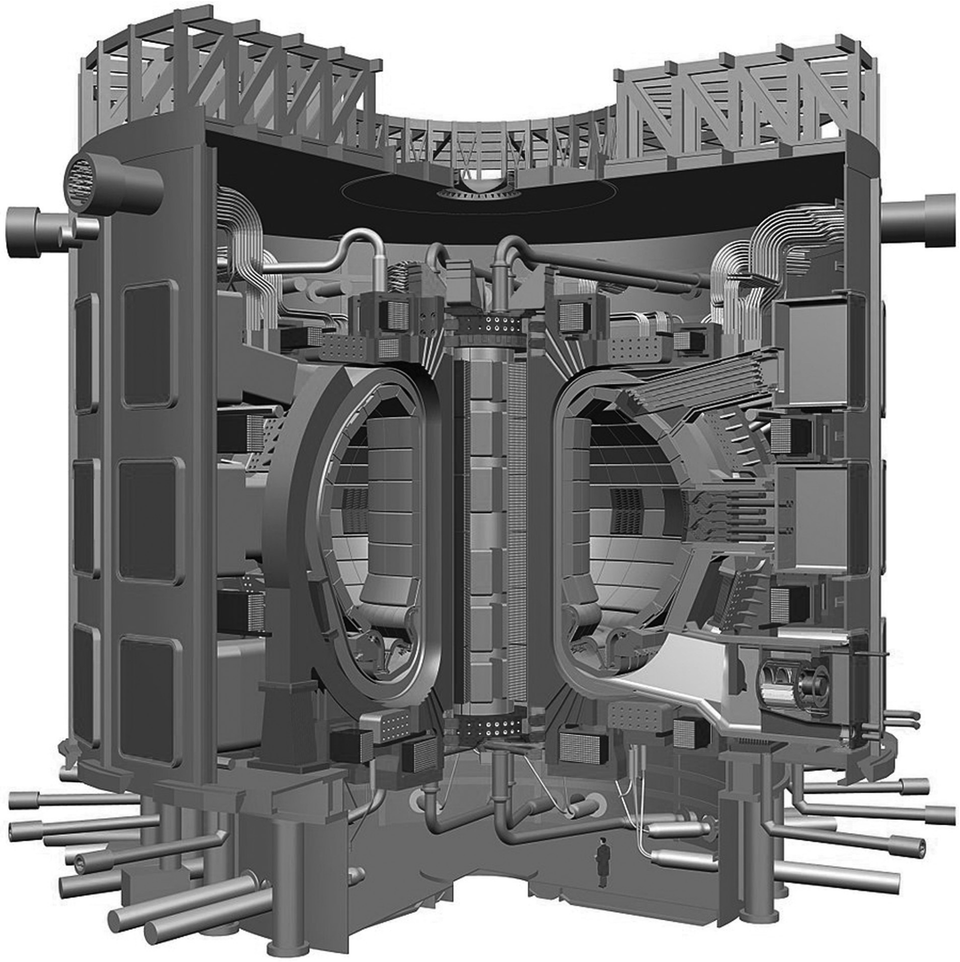

FIGURE 7.1: A tokamak. Credit: U.S. Department of Energy from United States/Public domain.
Source: https://commons.wikimedia.org/wiki/File:U.S._Department_of_Energy_Science_425_003_001_(9786811206).jpg

(see Figure 7.2) and “magnetic-mirror” devices. Fusion research initially focused on a variety of configurations; in the late 1960s, however, results from Russian research triggered a “stampede into tokamak technology” (Herman 1991, 96) that led to many groups focusing on this configuration. Expectations of reaching net-energy breakeven in the 1980s were not met, however: previously unknown instabilities were found to occur in toroidally confined plasmas. In response, plasma volumes and strength of magnetic confinement were both increased massively, up to the scale of ITER. Meanwhile, research into alternative configurations—especially stellerators—has experienced a modest revival. Compared to tokamaks, stellerators are steady-state machines without internal current. This eliminates some of the instabilities that have been found in tokamaks; however, stellerators require more powerful magnets for confinement.

A more specific illustration used in this chapter—the divertor—is related to one of the many problems of designing a functional reactor of any configuration. Fusion reactions inevitably produce "waste": ions that are too heavy to be reactants. Moreover, confined plasmas are bound to contain impurities from the vessel wall. Both waste and impurities may create instabilities, or otherwise interfere with the fusion process. The function of the divertor is to remove such material from the plasma, which is a far from trivial task. Waste-absorbing materials ("targets") in the divertor are exposed to heat fluxes estimated to be ten times higher than those to which space shuttles are exposed upon re-entering the Earth's atmosphere, and absorbing waste itself produces excess heat in the targets, which needs to be removed through cooling. Moreover, the divertor should absorb ions and impurities without creating new impurities and cooling the plasma; thus, simply inserting it into the outer layer of the plasma (where the heavier ions and impurities are located due to centrifugal effects) is counterproductive.

Designing a divertor involves a choice of suitable materials, of a configuration of the targets, and of a process through which waste is captured in the target, where these and many other complications need to be taken into account. For ITER, targets are made of tungsten; the divertor is located at the bottom of the reactor vessel; it consists of fifty-four ten-ton cassettes on a supporting structure mounted on rails; and absorption is achieved by diverting a small section of the plasma's outer layer (the "scrape-off layer") so that it touches the vessel wall where the cassette, containing a target, is located. With

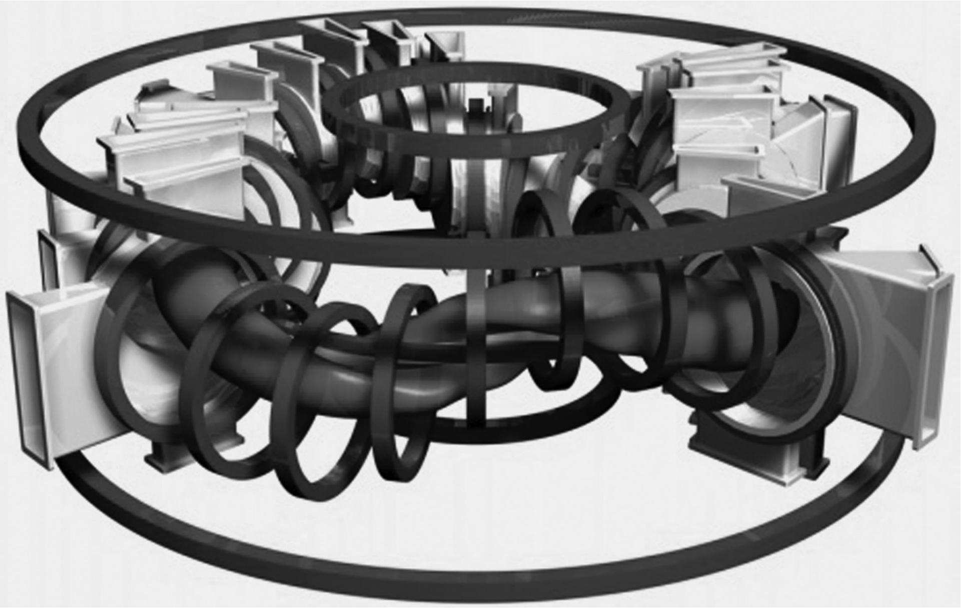

FIGURE 7.2: A stellerator. Credit: Wikimedia Commons/CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0)

Source: https://commons.wikimedia.org/wiki/File:TJ-II_model_including_plasma,_coils_and_vacuum_vessel.jpg this configuration, the divertor is supposed to last for ten years, so that it needs to be replaced only once in ITER's expected twenty-year operational lifetime.

One of ITER's seven "topical groups" is entirely devoted to developing and validating physical theories and models of the divertor and scrape-off layer. It studies, among other things, interactions between the plasma and candidate materials for the target, ways of storing absorbed ions, transport of ions and impurities, as well as instabilities in the plasma that may result in excessive heat loads and damage to the divertor. Such "edge-localized modes" (Leonard 2014) have long been known to occur, but the underlying physical mechanisms are not yet sufficiently understood. Here, we will take any knowledge that is produced in or directly useful for teams such as this topical group to be candidate engineering knowledge—since it is, apparently, of immediate relevance to some of the engineering challenges of divertor design.

Designing ITER or its divertor, including its placement, configuration and manner of operation, is not representative of all kinds of engineering, e.g., which involves standardized, mass-made consumer products in a highly competitive market. Rather, it is an extreme example of high-tech system design. This involves, roughly, archetypical "flagship" products of engineering—mid- to large-sized human-made goods, potentially one-off, of high complexity in terms of numbers of components, diversity of manufacturing and assembly processes involved, and interactions of parts. In Section 5, we consider to what extent our analysis in this chapter extends to epistemic aspects of other forms of engineering.

3. Existing Views of Engineering Knowledge

In this section, we offer a brief review of existing views on engineering as an epistemic activity.[1] These can be distinguished into three broad classes by how they conceive of the relation between engineering knowledge and the knowledge produced in the natural sciences. A first class takes this relation as one of *subordination*: if engineers produce any knowledge at all, it is by applying and specifying natural-scientific theories. A second class *contrasts* engineering knowledge to that produced in the natural sciences: engineering requires different, self-produced types of knowledge. Finally, a third class rejects (the need for) distinctions; rather, it *assimilates* science and engineering in practices of "technoscience."

3.1 Subordination

A first class of views echoes the (self-)characterization offered in campaigns such as "Born to Engineer." It holds that, epistemically, engineers primarily use, specify, or otherwise apply insights gained in the sciences—in particular, fundamental physical theories. Describing, for instance, the plasma-target interactions in a divertor or predicting the occurrence of edge-localized modes requires application of the basic principles of magnetohydrodynamics (MHD); it seems impossible, or remarkably ill-advised, to even try to design a magnetically confined plasma without this basic framework.

The most prominent subordination view is often attributed to Mario Bunge.[2] In his essay "Technology as Applied Science" (1966),[3] we find two versions: one that concerns *content* and another that concerns *method*. Content subordination mainly applies to "substantive technological theories," such as contributions to propeller theory or nuclear-reactor theory. These are said to be "essentially applications, to nearly real situations, of scientific theories" and as such, "always preceded by scientific theories" (1966, 331). Moreover, they are "invariably less deep" (1966, 333) since they are only concerned with controllable effects; and they may involve black-boxing aspects that are in principle captured by scientific theories (e.g., turbulence around propeller blades).

A chronological version of this view—engineering knowledge requires *prior* theorizing in physics—is easily dismissed. Often, scientific theorizing lags behind technological development and engineering knowledge about innovative technologies. Theories of thermodynamics and various other nineteenth-century scientific breakthroughs were directly inspired by the limited applicability of existing scientific theories to prior technological developments (see Channell 2009 for an overview). Density limits in tokamaks (reviewed in, e.g., Greenwald 2002) provide a more recent example. It has been established that exceeding such limits typically leads to disruptive instabilities. This is crucial knowledge in fusion engineering, which tempered initial optimism about producing fusion power with tokamaks and led to the vast upscaling of the ITER project—but the underlying physical mechanisms remain unclear.

Content subordination, alluded to in Bunge's "depth," concerns the epistemic merits of engineering knowledge, irrespective of chronology. It maintains that insofar as there is scientific knowledge about a topic, it is invariably epistemically superior to engineering knowledge—in having higher explanatory value, representational accuracy, generalizability, predictive power, and/or other epistemic virtue. Even if engineering knowledge would be prior to scientific theorizing, the latter would *improve* on the former. This presupposes commensurability of both types of knowledge, as well as substantial overlap in content—including use of the same basic concepts or governing principles and concern with the same domain.

Content subordination has not been developed in much detail, neither in Bunge's essay nor elsewhere. Still, the general thrust is familiar from the (history of) philosophy of science: it casts engineering knowledge as a "special science." Special sciences were once similarly held in low epistemic regard in comparison to fundamental physics by proponents of (specific versions of) the Unity of Science ideal. However, most have abandoned this ideal on the basis of arguments for the irreducibility and autonomy of the special sciences (see, e.g., Cat 2017) and against the supposed high epistemic value of law-like statements (see, e.g., Cartwright 1983). What might distinguish content subordination from unity-of-science claims more generally is that it is combined with claims about the aim of engineering practice: producing knowledge is not the ultimate aim, but rather instrumental—recall the "Born to Engineer" slogans. We shall discuss this claim in more detail later. For now, we note that, even if it were true, it does *not* support content subordination. A difference in overall aim might make engineering theories irreducible to theories in the natural sciences: Basic theoretical concepts may be homonyms (Kroes 1992) because terms such as "pressure" refer to physical characteristics in the natural sciences and to technical characteristics of designed objects in engineering.

Another, *methodological* form of subordination is found in Bunge's analysis of technological rules. Such a rule "[prescribes] the course of optimal practical action" (1966, 330), and is "grounded" if and only if it is "based on a set of law formulas capable of accounting for its effectiveness" (1966, 339). This identifies some central content of engineering knowledge as having a form distinct from that of scientific knowledge: where the latter is descriptive, the former is prescriptive or rule-like (see also Zwart 2019). Rather than expressing a contrast view, this may involve another kind of subordination. In Bunge's words: "in order to be able to judge whether a rule has any chance of being effective, as well as in order to improve the rule and eventually replace it by a more effective one, we must disclose the underlying law statements, if any" (Bunge 1966, 339). Here, proper grounding of rules is claimed to increase their *practical* value: without subordination of engineering knowledge to the sciences, it involves leaps of faith regarding effectiveness, or mere trial and error.

In its reference to law-like statements, this might appear a variant of content subordination: rules are held to be grounded through disclosing their relation to (paradigmatically law-like) physical theories. Yet the underlying law-like statements need not be identical to the contents of any scientific theory. Then, engineering knowledge may still be held subordinate to science in its *method*: science does not rest with stating brute empirical regularities, but aims at disclosing underlying law-like statements, thus improving predictive and explanatory power. Engineering should fully emulate this scientific method of postulating and validating hypotheses derived from general (law-like) theories.

Methodological subordination of engineering knowledge as sketched by Bunge is problematic—again, in part, because it imposes an image of scientific method that philosophers of science have found wanting. Other problems more specifically concern rule-based knowledge and its role in engineering practice. First, many rules are justified on the basis of prior experience or testimony: the recommendation not to exceed density limits in a tokamak is an effective way to avoid some instabilities, but it is not law-based. Second, the value of scientific knowledge in *improving* or *optimizing* practical courses of action is doubtful: insisting on this value appears to misunderstand what is involved in such optimization. Some courses of action—such as the shortest route to the railway station—can be optimized, but are too trivial to require law-like statements. Engineering problems frequently involve trade-offs between many (non-epistemic) values, such as efficiency, safety and reliability (e.g., Van de Poel 2009). Grounding as envisaged by Bunge does not resolve or even address such trade-offs: magnetohydrodynamics might offer fundamental, law-like statements for the behavior of plasmas, but it does not tell whether to opt for more magnetic power and fewer instabilities (as in stellerator designs) or, as in ITER, comparatively less magnetic power and more potential instabilities. We return to this aspect of engineering knowledge in Section 4.3.

3.2 Contrast Views

Subordination views have few champions and many outspoken critics. Perhaps the most radical alternative maintains that, as an epistemic activity, engineering is fundamentally different—and therefore autonomous—from the natural sciences. This *contrastive* view has been expressed since the 1960s in a number of places. Its context of development is not primarily that of responding to the descriptive and normative shortcomings of Bunge's applied-science thesis and linear models of innovation: it cannot be disengaged from attempts to emancipate engineering education—in particular its design elements. Two of the most influential expressions, offered in different contexts and for different audiences, are Nigel Cross's (1982) defense of specific "designerly" ways of knowing and Walter Vincenti's (1990) analysis of engineering knowledge, based on extensive research into aeronautical history.

Contrast views vary in their ambition level. Some seek to emancipate engineering and design completely from the natural sciences and research, others point out some continuities. A useful distinction is between *goal-contrasting* and *knowledge-contrasting* views, where the former are to some extent conceptually prior to the latter. Goal-contrasting views maintain that science and engineering have fundamentally different aims: the point of science is to describe or explain, that of engineering is another. Two statements to this effect are the following:

> "A natural science is a body of knowledge about some class of things, objects or phenomena in the world: about the characteristics and properties that they have; about how they behave and interact with each other . . . The engineer, and more generally the designer, is concerned with how things *ought* to be."
>
> (Simon 1981, 3, 7)

> "Science concerns itself with what is, technology with what is to be."
>
> (Skolimowski 1972, 44)

As intuitively plausible as such goal-contrasting statements are, they suffer from a number of problems as analyses of engineering knowledge. First, as is revealed by the citations, it is unclear what is contrasted with science: engineering, or design—where design can both be taken as more general, including for instance curriculum design or therapeutic design, and more specific, excluding engineering *science*. Second, although it is relatively clearly stated what the aim of science is, the aim of engineering/design/technology is less clear. Ethics and many religions are also concerned with how things ought to be, and futurology and science fiction with what is to be, but neither is a branch of engineering. Third, goal-contrasting views are not necessarily views of engineering *knowledge*: they are compatible with views on which knowledge production is not an aim of engineering at all. Such views offer a clear-cut contrast: whereas science should produce knowledge to explain, describe, or predict, engineering should result in functional devices. But this contrast comes at the price of ignoring any epistemic impact of engineering. Fourth, even if goal-contrasting views would concern engineering knowledge, they might collapse into subordination views. Bunge's technological rules concern how things ought to be and how to make them so. However, as argued earlier, they do not support a contrast between science and engineering, or at least not one that is conducive to the epistemic status of engineering.

Many advocates of the epistemic autonomy of engineering go beyond goal-contrasting to insist on *knowledge*-contrasting. This contrast does not merely concern the content of engineering knowledge: such differences are also found within the sciences, where fields are routinely distinguished in terms of their subject matter (e.g., condensed-matter physics, astrophysics). Knowledge-contrasting views are considerably more ambitious in insisting that "technological praxis [is] a *form* of knowledge" (Staudenmeier 1985, 120), that there are "designerly *ways of knowing*," or that technology is "an autonomous body of knowledge, identifiably different from the scientific knowledge with which it interacts" (1990, 1–2).

The idea is that engineering knowledge is different in its epistemology (e.g., as different as testimonial knowledge is from observational knowledge). Cashing this out is difficult, and knowledge-contrasting views are understandably divergent and somewhat provisionally formulated. Often, they take the form of listing some characteristic elements of (paradigmatic) engineering knowledge or of attempting a comprehensive taxonomy (see Houkes 2009, Section 4 for an overview of such attempts). Some elements that have been proposed in multiple knowledge-contrasting views are

1. Prescriptive: engineering knowledge concerns what ought to be done and/or made. This knowledge connects human needs to their artificial environment, prescribing ways of changing this environment or ways in which humans can interact with it. By contrast, scientific knowledge describes reality, including underlying mechanisms or future states of the world.
2. Tacit: the nature of design problems and solutions is such that engineering knowledge cannot always or typically be made fully explicit, but that it is tacit, implicit "knowing-how." By contrast, scientific knowledge is fully explicit, propositional, "knowing-that."
3. Embodied in objects: engineering knowledge may in part "reside" in products of engineering activities that are not themselves epistemic. By contrast, scientific knowledge primarily resides in theories and other propositional content.

There are several general problems with these characteristics and the views that appeal to them. First, perhaps precisely because they do not apply to traditional forms of knowledge, all characteristics tend to be under-analyzed. Consequently, they may be indeterminate or ambiguous (Nightingale 2009); or, on closer analysis, reducible to traditional forms of knowledge (Stanley and Williamson 2001). Analyses that hint at a contrast and apply an evocative label to the engineering side of this contrast do little to clarify engineering knowledge.[4]

Second, contrastive views not only need to state what engineering knowledge is, but also what—by contrast—*scientific* knowledge is. Implicitly or explicitly, contrastive views might rely on controversial views of science, such as naïve forms of scientific realism. In addition, contrastive views need to elevate some contrasting characteristics to demarcative status. They might therefore overlook or downplay continuities or aspects of practices on either side of the established contrast. In many sciences, for instance, knowledge resides in things; think of experimental equipment, model organisms, and scale models (Baird 2004). One might still insist that engineering knowledge does so to a larger extent, but that would undercut arguments that it constitutes another *form* of knowledge altogether; or that it does so in a different way, but that would beg the question. Here, one might wonder why the (important) project of providing an autonomous analysis of engineering knowledge has turned into the (possibly misguided) project of providing an analysis of its autonomy.

Third, arguments that whatever results from engineering as an epistemic activity is knowledge of a special type are vulnerable to the objection that such results may certainly be valuable, but that they are not *knowledge*—precisely because they do not fit the traditional form. Prescriptive knowledge may, for instance, be argued to be of a special type (Zwart 2019) because claims to this knowledge are not candidates for being true (Von Wright 1963)[5]—but it might be objected to this that, then, it can only be called "knowledge" in an extended (Meijers and Kroes 2013) sense on any analysis that takes truth to be a necessary condition for being knowledge.

Fourth and finally, contrastive views might be of limited applicability. They might fit practices that are historically or intuitively distant from typical scientific practices, such as early episodes in aeronautical engineering, or practices in industrial design. However, whatever knowledge is produced by ITER's Scrape-Off and Divertor Topical Group is in most respects difficult to distinguish from that produced by experimental-physics groups; measuring and modelling heat loads in plasma facing components—to give one example of knowledge that may be produced in the group—are not more tacit or embodied in objects, or exclusively prescriptive than measuring, say, heat flows through condensed matter independently of any context of application. More importantly, there appears to be little point in insisting on some contrast: the association with experimental physics is also sufficiently close in terms of status. Practices such as divertor design do not need to be emancipated in order to be taken seriously as epistemic activities.

3.3 Assimilation Views

Assimilation views highlight some of the similarities between scientific and engineering knowledge that also featured in subordination views. However, they do not conjoin them with priority claims, or normative ideals about fundamental scientific theories. Instead, they conclude that, in historical and/or contemporary practice, science and engineering are so closely intertwined that it would be misleading to prize them apart conceptually.

Indeed, some advocates of assimilation prefer to refer to a broad suite of scientific and engineering activities by a single term: "technoscience." This term was used cursorily as early as the 1950s, and popularized by Gilbert Hottois in the late 1970s. It found the most influential use in Latour's (1987) groundbreaking study of laboratory life. Since then it has met with criticism, but still sees widespread use by researchers in science and technology studies, post-phenomenology, and cultural and gender studies.[6] For instance, in the subject index of the fourth edition of the *Handbook of Science and Technology Studies* (Felt et al., 2017), "technoscience" takes up an entire page—more than "science," and "technology" and "engineering" do not even feature.

In studies that employ this term, as well as more broadly in the fields just mentioned, assimilation views are one element of complex outlooks on the ontology, epistemology, and socio-political context of science and engineering.[7] Necessarily somewhat schematically, the following assimilation views of engineering knowledge may be distinguished:

a. *Association* (e.g., Hughes 1986; Latour 1987): practices of "science in the making" and "engineering in the making" involve irreducibly complex and close associations (networks) between theories, instruments, and a host of other actants.b. *Reversal* (e.g., Lyotard 1984, Forman 2007): technology and its capitalist mode of production have become the main drivers of scientific research.c. *Application* (e.g., Gibbons et al., 1994; Carrier and Nordmann 2010): scientific knowledge is increasingly produced with an eye to its applicability, so much so that research and design activities are indistinguishable or inextricably combined.

These views are in part distinct in their historical claims. Strong associations between science and engineering and/or technology have been around for centuries. By contrast, "reversalists" see a more or less sharp historical discontinuity in science around and after the Second World War, and especially in postmodern societies. Finally, "applicationists" find their strongest illustrations in even more recent cases, of converging technologies and corresponding research into such areas as nanotechnology, smart materials, and biomimetic design. This suggests that, as credible as these views are for the individual contexts and cases for which they have been developed, none of them may fully capture—diachronically or synchronically—the interrelations and distinctions between scientific and engineering knowledge; nor can they be straightforwardly regarded as complementary, given their conflicting historical claims.

Assimilation views are also distinct in their underlying theoretical frameworks, of which only a selection can be mentioned here. Latour's views are couched in actor-network theory (e.g., Latour 2005). Here, associationism is developed through the principle of "generalized symmetry," which holds that all entities in a network are in the first instance to be described in the same terms, and differences may only emerge in their network of relations. "Reversalists" rather appeal to (philosophical) postmodernism, which emphasizes the erosion of grand narratives since the end of the nineteenth century, including narratives of the primacy of science and its legitimization by modernist philosophy; and which offers means to expose and to some extent avoid such narratives. Finally, some highly influential theories have been proposed in support of applicationist views, such as the distinction between traditional “Mode 1” knowledge production and contemporary “Mode 2” production (Gibbons et al. 1994), or Triple Helix views (Etzkowitz and Leydesdorff 2000). These theories, which come in many varieties themselves, conceptualize fundamental shifts or “reconfigurations” in knowledge production, in terms of the organization of academic research, funding streams, and alignment of research agendas with technological innovation and societal challenges.

This conceptual diversity comes with a broad variety in normative implications (if any). Only occasionally, these implications concern the knowledge that is produced in these practices: some by-reversal claims lament the transition to “impure” science, but they offer no countermeasures and have been criticized as modernist nostalgia (Forman 2007). More often, the implications concern studies *of* scientific and engineering practices, namely to refrain from using (some) value-laden terms, sharp dichotomies or outdated “modernist” ideologies. Applicationists for instance insist that studies of contemporary practices such as synthetic biology should not impose ideas about the interrelation between research and design or about “proper” forms of knowledge production in science and engineering, but describe and assess practices in their own right.

This lack of implications for (contemporary) knowledge production has been criticized as implicit support for or for the emergence of transdisciplinary, application-oriented research and the commodification of knowledge (Godin 1998; Hessels and Van Lente 2008; Mirowski and Sent 2008). It also results in a lack of clear and detailed implications for knowledge practitioners. Subordination views offer guidelines, albeit misleading ones, such as requirements to ground rule-based engineering knowledge in law-like statements. Assimilation views mainly warn against imposition of explicitly normative frameworks, and describe how knowledge production actually works, especially on an institutional level: they focus on the context, rather than the content, of epistemic activities (Houkes 2016). As a quick illustration: assimilation views could be applied to ITER, or perhaps more specifically to divertor design, to reveal how traditional research and design activities have merged here; or how it exemplifies contemporary transnational projects and funding streams, perhaps in contrast to more traditional “Big Science” projects such as the Large Hadron Collider. As such, it does not concern any specific knowledge claims produced in the context of this project, or their relation to other claims and epistemic activities: unlike contrast views, assimilation views do not seek to distinguish, say, outcomes about workable divertor configurations from the basic equations of magnetohydrodynamics. Unlike subordination views, they do not insist on some (formalizable) relation between such outcomes and fundamental theories.

4. Ingredients for an Alternative Analysis

In this section, we offer several ingredients for an analysis of engineering as an epistemic enterprise, focusing on high-tech systems design. We bring out how it involves sets of *epistemic activities*, resulting in a variety of *rules*, which are governed by a distinctive set of epistemic and *non-epistemic values*. The ingredients are largely based on recent work in the philosophy of science.

4.1 Practices and Epistemic Activities

Contemporary philosophy of science is largely unconcerned with demarcation issues or formal reconstructions. Rather, many in the field take pluralistic, often practice-based perspectives. Formal reconstructions of law-based explanations in physics, with strong normative aspirations for other disciplines to follow suit, have been exchanged for more detailed and in-depth presentations of experimentation, simulation, modelling, data collection and other activities in the natural, life, and behavioral sciences, without strong normative implications beyond identifying the diversity of scientific best practices. One attempt, in this context, to characterize epistemic practices is found in recent work by Hasok Chang (2011, 2012). Chang distinguishes several hierarchically ordered units of analysis, which may be distinguished depending on the context:

- *Scientific practices*, which have characteristic aims and consist of more or less coherent sets of- *Epistemic activities*, which are intended to contribute to knowledge production in accordance with discernible rules, and consist of more or less coherent, routinized sets of- *Mental and physical operations*

Lavoisier's revolutionary way of doing chemistry, for example, can be understood as a new practice in its time, with activities such as collecting gases, classifying compounds, and measuring weights. These are all epistemic in the context of being performed in a scientific practice. Chang's analysis does not offer a characterization in terms of aim alone: it brings out how differences in aim might affect the constitutive level of activities. One such effect may be that practices with different aims comprise different, but overlapping sets of activities. Another effect, more difficult to pinpoint in realistic cases, is that because of differences in aim, the same activities "fit together" differently. Coherence, at least at the level of activities, is a matter of effective coordination with regard to the overall aim of the practice constituted by the activities. Thus, we might see divergent standards of quality or differences in integration with other activities (e.g., when, where, and who performs the activity).

Take, for instance, an activity that is both enabled by and involved in high-tech system design: investigating thin films of materials by high-resolution x-ray diffractometry (HRXRD). Narrowly speaking, this activity serves the (epistemic) purpose of obtaining information about the structure of a layered film, such as its roughness, thickness and density. This activity will fit together differently with other activities depending on the encompassing practice: HRXRD is used in the semiconductor industry to, among other things, analyze defects in multilayered devices and establish whether such defects are caused by, for instance, faults in stacking layers. In industrial research, one might rather investigate how HRXRD produces scanning results for different crystalline structures and establish its limitations in analyzing specific structures. In biomedical research, the technique may be used to find out the crystalline structure or composition of living bone and various substituent materials, in order to establish their biocompatibility (e.g., Peters et al. 2000). In archaeology, researchers may use HRXRD to investigate the composition of ancient glass, for instance to obtain information about melting processes used in producing it (e.g., Janssens 2013, chapter 2.1).

All practices arguably involve the same activity, and some operations may be similarly routinized or may be governed by the same rules. For instance, in any application, the sample to be scanned should be mounted properly; bending or scratching is to be avoided since this would obviously interfere with whatever results are to be obtained. Still, the different practices may lead to differences in how and when the activity is performed. For instance, in the industrial-research practice, the activity is performed to analyze *known* features; in the other practices, to analyze *unknown* features. In each of the mentioned practices, different features of the scanned material are relevant. In the industrial practice, HRXRD might be applied repeatedly as part of quality control in a manufacturing process; in the biomedical-research practice, quality control might also be the ultimate aim, but it may be a one-off process that results in rejection of some materials as insufficiently biocompatible.

Another effect of contextual differences is in how the practice deals with limitations or uncertainties of the technique. As a diffraction technique that relies on analyzing reflections from multiple angles, HRXRD has limited detection depth and scans may be blurred by surface imperfections. These limitations might not affect diagnostic applications in the semiconductor industry, as long as layered films are sufficiently thin. In biomedical or archaeological research, the same limitations and uncertainties might require use of supplementary analytic techniques. These differences might not only be a result of variation in the properties of the sample of interest (e.g., thin films or ancient glass), but also of the implications of incorrect information—the *inductive risk* (Douglas 2000) that is run by performance of the activity (see Section 4.3). Drawing the wrong conclusion, or failing to draw the right conclusion, about the melting process used in producing 14th century Andalusian glass does not do the same harm as drawing the wrong conclusion about substitutes for living bone; and the probabilities of false negatives, false positives or both are also likely to be different in these applications.

Using Chang's grammar of practices and activities may reveal how (or even whether) epistemic practices in the sciences and in various engineering contexts may involve different constitutive activities, or how they differ in integrating the same activities, harmonizing them with other activities, or dealing with their limitations. This would, to some extent, develop Bunge's *methodological* form of subordination: engineering practices might involve activities that originated in scientific practices, such as using mathematical models to derive predictions, to give just one broadly defined example. Conversely, however, scientific practices may involve engineering activities—in line with (associationist or applicationist) assimilation views.

This being said, analyses of activities within practices are unlikely to reveal a *uniform* difference between science and engineering. It is, to return to the example, not to be expected that using HRXRD has, as an epistemic activity, the same role in every engineering practice: there may be significant differences between its use in, say, materials science to investigate nanostructures and in software engineering to enhance data analysis. Shared features—if any—may emerge only from in-depth analyses of engineering practices. We take this as a strong advantage rather than a drawback.

4.2 Design Rules

An analysis on the level of practices and activities can be supplemented at the level of individual knowledge claims. Here—in line with subordination and contrast views—we focus on *prescriptive* or *rule-based* statements, without claiming that this is distinctive for engineering practices, or needs to be grounded in law-like statements. More specifically, we develop an account proposed by Torsten Wilholt (2006), who identifies some of the epistemic products of "industrial research" as "design rules," which have the form:

> (DR) "*A* → *B*, where *B* . . . describes one or more properties of a system that are interesting for its applications, and *A* describes a set of characteristics of the same system that can be controlled during its production."
>
> (Wilholt 2006, 79)

One example given by Wilholt concerns spin valves, which contain layers of thin films of various materials. If layers have the right materials, thickness, etc. (characteristics *A*), their magnetization directions can be highly sensitive to external magnetic fields, which makes spin valves interesting as sensors (characteristics *B*).

Like Bunge, Wilholt brings out the value of research that aims at such prescriptive knowledge as enhancing our ability to realize goals *effectively* and *efficiently*. However, Wilholt avoids Bunge's appeal to law-like statements. Rather, he shows how successful industrial research may be guided by *models* of a device. These allow the identification of promising specifications, that is, propositions A and/or B in (DR), without extensive trial-and-error. Models of magnetization effects in spin valves, for instance, narrow down the choice of materials and thicknesses, drastically reducing the number of possible configurations to be tested for the desirable properties. Based on Wilholt's proposal, several knowledge claims may be distinguished, although they are often combined in practice. (DR) conjoins two types of statements:

> Manufacturing rule (MR): "If a set of characteristics *A* of a system is controlled during its production, it will have a specific set of properties *F*"
>
> Functional knowledge (FK): "If a system has a set of properties *F*, it will (usually) be able to perform the function to *Φ*"

Thus, in Wilholt's example, (MR) concerns properties such as thickness and materials that can be directly controlled during manufacturing (*A*), and that ensure that "magnetostatic coupling and ferromagnetic interlayer coupling cancel each other out" (2006, 80). This property *F* in turn explains why the resulting spin valve has the disposition of being highly sensitive to external magnetic fields, which in particular contexts of applications may be highlighted as its function (Houkes and Vermaas 2010). A variant of (MR) that is more process-oriented and may similarly be combined with functional knowledge, is

> Operational rule (OR): "If a set of characteristics *A* of a system is controlled during its operation, it will have a specific set of properties *F*"

Thus, operation of a particular type of spin valve may require constant temperature, or require sudden changes in temperature to manifest the desired disposition.

Epistemic activities in engineering practices might result in knowledge of one type of rule without the other.[8] It may be known how to control for the characteristics of a device or material during production, but unknown why those characteristics give rise to desirable behavior—or to what extent it may be avoided that they show undesirable behavior. Plasmas and their instabilities are cases in point: it is known that they arise, and how they can to some extent be avoided, but edge-localized modes (ELMs) are, for instance, insufficiently well-understood to estimate their effects on target plates in divertors. Hence, ITER's Topical Group must produce sufficient insight into ELMs to avoid them during operation—allowing statement of operational rules for, for instance, the magnetic confinement of plasmas—or to design divertors such that they can withstand the effects of repeated ELMs—allowing statement of manufacturing rules for divertors. In the absence of such insight, extensive trial and error is required to realize or optimize useful devices, which is unfeasible on the scale of ITER.

Wilholt makes clear how rule-based knowledge in industrial research may be generated through modelling the system to be designed. This holds more generally for engineering knowledge,[9] also in high-tech systems design, and brings out an interconnection with the first ingredient: a variety of epistemic activities has been developed in engineering contexts in order to search design spaces more quickly than by trial and error. Such activities may, on some level of description, be the same as in scientific practices (e.g., "modelling")—and this is useful for highlighting continuities, as in assimilation views. On another level of description, however, the difference in outcome (descriptive/explanatory statements versus prescriptive, rule-like statements) is bound to be reflected in subtle differences in epistemic activities. Some may manifest in different forms of inference (e.g., Zwart 2019), others in different standards of acceptance, still others in different focal points. Modelling physical infrastructures in scientific and in engineering practices may, for instance, both involve the discovery of mechanisms (Craver and Darden 2013). Yet the latter may highlight triggering conditions of mechanisms that are or can be directly related to human interventions, leading to forms of mechanistic explanation (e.g., Van Eck 2015) or trajectories of developing mechanism schemas (Houkes 2016) that are characteristic for (some) engineering practices.

4.3 The Role of Non-epistemic Values

A third and final ingredient of our analysis concerns the broader context of epistemic activities in engineering practice and the role of non-epistemic values. Traditionally, philosophers have maintained that science ought to be value-free: apart from, possibly, the choice of topics, ethical restrictions on methods and the presentation and further application of results, influences of non-epistemic values (reflecting societal, commercial and political interests) are illegitimate and compromise the epistemic merits of research outcomes. Recently, several powerful arguments have been offered against this value-free ideal.

One family of arguments[10] centers on "inductive risk." Scientific reasoning is fallible: scientists may draw incorrect conclusions, or fail to draw the correct conclusion. There are several aspects to this risk, such as deciding how much evidence is sufficient to state a conclusion, and determining what counts as evidence in the first place. Now, in many cases, choosing *any* standard of evidence, for instance concerning a human-induced greenhouse effect, has serious societal repercussions, since it affects the balance of over- or underestimating the effect and thus over- or under-regulating its purported causes. If scientists can reasonably foresee these consequences, they ought to account for them and mitigate inductive risk. Thus, choosing standards for the quality and quantity of evidence *ought* to be influenced by the non-epistemic values with which societal consequences are assessed: a toxicologist may legitimately use low standards of evidential quality (e.g., in assessing borderline cases) if she considers under-regulation of some substances to be harmful.

A brief investigation suffices to reveal how broadly and deeply inductive-risk arguments apply to engineering practice. Efficiency and effectiveness play prominent parts in engineering-design methodologies. However, the role of societal values such as safety, sustainability or commercial viability is equally undeniable. As alluded to in Section 3.1, the problems faced by engineers are characterized by trade-offs between multiple values. The choice of a refrigerator coolant, for instance, is governed by non-epistemic values such as toxicity, flammability, and atmospheric lifetime, as well as instrumental values such as the cost and scalability of production—which cannot each be optimized simultaneously (Van de Poel 2009). Likewise, the "design space" of spin valves (Wilholt 2006), of photovoltaic materials (Houkes 2016), and of fusion reactors and tokamak divertors does not have a single peak, but is an extremely rugged landscape, with a very large number of local optima (see Figure 7.3). The individual optima are not just hard to find in a given landscape: the trade-off between multiple values and even selecting which values are at play are such that engineering problems require continuous (and necessarily contentious) reconstruction of the landscape.

As a brief example, consider organic photovoltaics. Compared to the dominant silicon-based devices, photovoltaic cells that use polymers are more lightweight, flexible, and easy to produce; but they are currently far less efficient and stable. Consequently, there is great potential in research into the properties of organic

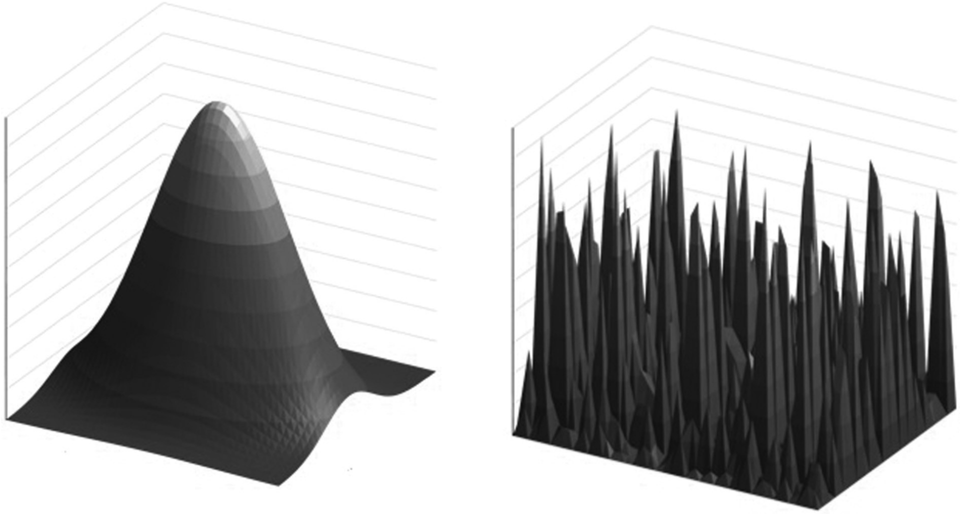

FIGURE 7.3: Single-peak and rugged search spaces. Credit: Thomas Shafee/CC BY (https://creativecommons.org/licenses/by/4.0)

Source: Adapted from https://commons.wikimedia.org/wiki/File:Epistasis_and_landscapes.png

photovoltaic materials and ways of producing them: generating design rules of the various forms discussed in Section 4.2 is a valuable epistemic engineering practice. However, there are great opportunity costs as well, in terms of time and effort not invested in further optimizing silicon-based devices. Knowledge claims about any of these properties of a material, the devices in which it is used, and the manufacturing process carry substantial inductive risk, because it is never fully clear whether further improvements in one of the many performance characteristics are possible and technically feasible: the landscape is not only rugged, but also foggy. The search for a local optimum by considering a particular type of material can be affected dramatically by relevant new knowledge. It has, for instance, been known for decades that perovskite materials (organic-inorganic compounds) have features that might make them suitable for applications in photovoltaics, but only around 2010 spectacular improvements were made for perovskite photovoltaics in terms of efficiency. This then led to a surge of research into ways of not just making further improvements, but also of overcoming major performance issues for these types of materials, such as their degradability in moist air and lead content: without promises of outperforming other photovoltaic materials in terms of efficiency, these performance issues would not have been worth investigating.

This “perovskite gold rush” may seem similar to the “tokamak stampede” mentioned in Section 2. However, a closer comparison brings out other ways in which non-epistemic values impinge on high-tech systems design. Searching the rugged, foggy landscape of photovoltaic materials through various epistemic activities is a costly undertaking. Yet the costs of roads taken and not-taken are marginal in comparison to those in fusion engineering, especially at the scale needed for net-energy production; these require a much more selective search strategy. The tokamak stampede is therefore much riskier behavior[11] than the perovskite gold rush—and (re-)diversification to alternative configurations is a rational response to the upscaling required for tokamaks, which further increases the risks of the epistemic activities in fusion engineering leading to negative results after a lengthy and costly search.

These broader repercussions, and consequent inductive risk, are easily overlooked in a focus on design rules and functional knowledge, or even on epistemic activities. Stating a rule suggests that a particular course of action is worth taking: an item or its properties are claimed to be *sufficiently* interesting to consider actual manufacturing or at least to warrant further research. Whether this claim holds true is, however, highly sensitive to the practical context, including alternative ways of achieving the same goal or known practical difficulties.

This can be illustrated through the design choices for ITER's divertor. Suppose that ITER's Topical Group comes up with a particular manufacturing or operational rule for reducing the effects of ELMs on the divertor's targets. Such rules are hardly generalizable to other fusion-reactor configurations: in order to make sense for ITER, they need to take into account several local features. This makes epistemic activities for generating this rule-based knowledge valuable in the context of ITER; but investing in these activities also raises the stakes of fusion engineering in tokamak configurations—and therefore increases inductive risk. The value may even be restricted to ITER alone, because of the choice of material for the divertor targets. Tungsten, chosen because of its high melting point, may be suitable for the conditions in ITER, but because of its brittleness it is not necessarily the most suitable candidate for the planned *next* generation of fusion reactors—which will "open the way to industrial and commercial exploitation" (ITER Organization 2019). Epistemic activities that result in useful rules for tungsten-based divertors are valuable in their present context—and as results of "experimental tools," they may be perfectly valid. However, in the broader context, it is doubtful whether these rules and generative epistemic activities are worth investing in exclusively. Also here, diversification partly overcomes the problem: alternative divertor materials and configurations are being tested in smaller research facilities. However, scalability of results is such a common issue in fusion research that diversification would only be fully effective if implemented on equal scales—which is technically (and financially) infeasible.

5. Conclusion and Outlook

In this chapter, we have reviewed three types of existing views on engineering knowledge—subordination, contrast, and assimilation views—and we have discussed some of their descriptive and normative shortcomings. These include reliance on inadequate theories of science (subordination), inability to capture the specifically *epistemic* aspect of engineering practices (contrast), lack of guidance of such practices (assimilation), and a tendency to overgeneralize from specific cases to general views (all views).

We have also sketched three ingredients of an alternative view, which incorporates elements of all existing views: from assimilation views, we adopt a focus on epistemic activities within practices—and without a predetermined difference between scientific and engineering practices, or a predetermined similarity between all engineering practices; from subordination and contrast views, we adopt a focus on rule-based engineering knowledge, without claims regarding reducibility to or grounding in scientific knowledge; and from contrast and assimilation views, we adopt a focus on the importance of non-epistemic values in the development of engineering knowledge.

These three ingredients concern, roughly, different levels of analysis: a macro-level of non-epistemic values related to societal challenges or (potential) applications; a meso-level of epistemic and other activities through which such applications are produced; and a micro-level of epistemic products of such activities. Analysis of actual cases requires all three levels; the shortcomings of existing views may be partly due to excessive focus on one of these levels, at the expense of others. We have, however, only hinted at how these levels are interrelated, and at how their interrelations should be taken into account when studying engineering practice. Moreover, each individual ingredient requires further processing—both the discussion of existing views and the sketch of our own alternative show that analyses of engineering knowledge have hardly reached the level of sophistication of analyses of scientific knowledge (at least as offered in philosophy of science). This may, in itself, be one of the most unfortunate side-effects of all existing views, which may be mitigated by applying insights from contemporary, practice-oriented philosophy of science to improve our understanding of engineering knowledge.

Another point for further research concerns the scope and depth of our alternative. We have indicated how it accounts for epistemic aspects of one type of engineering practice: high-tech system design. There, we focused on features of one case—ITER—and some other illustrative examples. Analysis of further details, other features and different cases of high-tech system design is likely to require development and modification of our ingredients, or supplementing them with others. Finally, in focusing on one type of engineering practice only, we mean to avoid the risk of overgeneralizing our insights to all engineering knowledge. To paraphrase ourselves in closing: analyses of non-epistemic values, epistemic activities, and rule-based knowledge are unlikely to reveal *uniform* differences between science and engineering. Likewise, shared features—if any—may emerge only from in-depth analyses of the epistemic aspects of engineering practices.

Acknowledgments

We thank the editor of this volume, members of the Philosophy & Ethics department at Eindhoven, and especially Sven Ove Hansson for helpful feedback on earlier drafts of this chapter.

Notes

1. See Kant and Kerr (2019) for a more detailed and historically informed review.
2. Other advocates of subordination views may include the researchers of the *Starnberger Schule* (Böhme et al. 1976). They argued that research in scientific fields may legitimately be guided by external (e.g., commercial) goals once fundamental theories in these fields are "closed," i.e., there are explanatory laws with sufficient predictive power that cover the field's subject matter. Association of this line of work with some of the more maligned aspects of subordination views is even harder than it is for Bunge's essay. See Radder (2009, Section 4) and Houkes (2016) for further discussion.
3. In a later essay, Bunge (1988) offers a substantially different view of the relation between science and technology.
4. Analyses of engineering knowledge that go beyond the contrastive and build on these elements can be highly informative. A case in point is Hansson's (2013) characterization, which combines a typology—including both tacit and prescriptive knowledge—with an account of transformations between knowledge types.
5. See Niiniluoto (1993) for a defense of prescriptive knowledge claims as candidates for truth.
6. For a more detailed presentation of the history and critical reception of the term, see Channell 2017; Bensaude-Vincent and Loeve 2018.
7. Other elements are alternatives to the linear model of innovation and constructivist ontologies of scientific facts and technological objects.
8. Neither manufacturing/operational rules nor functional knowledge are exclusive to engineering. Biologists and cognitive scientists also deal in the latter; and developmental explanations in the life sciences may have a form similar to (MR).
9. Norström (2011) makes a powerful case for "know-how from rules of thumb": rule-based knowledge in engineering that is *not* based on idealized models.
10. Inductive-risk arguments were recently revived by Douglas (2000) and have a rich history, going back at least to Rudner (1953) and Hempel's (1960) defense of the value-free ideal of science.
11. This ignores the opportunity costs of investing in fusion power rather than other sources of renewable energy, such as photovoltaics. This shows that identification of the relevant design space (fusion energy versus renewable energy) is contentious in itself.

References

Baird, Davis. 2004. *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley: University of California Press.

Bensaude-Vincent, Bernadette, and Sacha Loeve. 2018. "Toward a Philosophy of Technosciences." In *French Philosophy of Technology. Classical Readings and Contemporary Approaches*, edited by Sacha Loeve, Xavier Guchet, Bernadette Bensaude-Vincent, 169–186. Dordrecht: Springer.

Böhme, Gernot, Wolfgang van den Daele, and Wolfgang Krohn. 1976. "Finalization of Science." *Social Science Information* 15: 307–330.

Born to Engineer. 2019. https://www.borntoengineer.com/

Bunge, Mario. 1966. "Technology as Applied Science." *Technology & Culture* 7: 329–347.

Bunge, Mario. 1988. “The Nature of Applied Science and Technology.” *Philosophy & Culture* 2: 599–604.

Carrier, Martin, and Alfred Nordmann, eds. 2010. *Science in the Context of Application*. *Boston Studies in the Philosophy of Science* (vol. 274). Dordrecht: Springer.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Cat, Jordi. 2017. “The Unity of Science,” *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), edited by Edward N. Zalta. Palo Alto, CA: Stanford University. Retrieved from: https://plato.stanford.edu/archives/fall2017/entries/scientific-unity/

Chang, Hasok. 2011. “The Philosophical Grammar of Scientific Practice.” *International Studies in the Philosophy of Science* 25: 205–221.

Chang, Hasok. 2012. *Is Water H_2O?* Dordrecht: Springer.

Channell, David F. 2009. “The Emergence of the Engineering Sciences.” In *Handbook of Philosophy of Technology and Engineering Sciences*, edited by Anthonie W.M. Meijers, 117–154). Amsterdam: Elsevier.

Channell, David F. 2017. *A History of Technoscience*. London: Routledge.

Craver, Carl F., and Lindley Darden. 2013. *In Search of Mechanisms*. Chicago: The University of Chicago Press.

Cross, Nigel. 1982. “Designerly Ways of Knowing.” *Design Studies* 3: 221–227.

Douglas, Heather. 2000. “Inductive Risk and Values in Science.” *Philosophy of Science* 67: 559–579.

Etzkowitz, Henry, and Loet Leydesdorff. 2000. “The Dynamics of Innovation.” *Research Policy* 29: 109–123.

Felt, Ulrike, Rayvon Fouché, Clark A. Miller, and Laurel Smith-Doerr, eds. 2017. *The Handbook of Science and Technology Studies*. 4th ed. Cambridge, MA: MIT Press

Forman, Paul. 2007. “The Primacy of Science in modernity, of Technology in Postmodernity, and of Ideology in the History of Technology.” *History and Technology* 23: 1–152.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow. 1994. *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*. London: Sage.

Godin, Benoît. 1998. “Writing Performative History?” *Social Studies of Science*, 28: 465–483

Greenwald, Martin. 2002. “Density Limits in Toroidal Plasmas.” *Plasma Physics and Controlled Fusion*, 44: R27.

Hansson, Sven Ove. 2013. “What Is Technological Knowledge?” In *Technology Teachers as Researchers*, edited by Inga-Britt Skogh and Marc J. de Vries, 17–31. Dordrecht: Springer.

Hempel, Carl G. 1960. “Inductive Inconsistencies.” *Synthese* 12: 439–469.

Herman, Robin. 1991. *Fusion: The Search for Endless Energy*. Cambridge: Cambridge University Press.

Hessels, Laurens K., and Harro van Lente. 2008. “Re-thinking New Knowledge Production.” *Research Policy*, 37: 740–760.

Houkes, Wybo. 2009. “The Nature of Technological Knowledge.” In *Handbook of Philosophy of Technology and Engineering Sciences*, edited by Anthonie W.M. Meijers, 309–350. Amsterdam: Elsevier.

Houkes, Wybo. 2016. “Perovskite Philosophy.” In *Philosophy of Technology after the Empirical Turn*, edited by Maarten Franssen, Pieter E. Vermaas, Peter Kroes, and Anthonie W. M. Meijers, 195–218. Dordrecht: Springer.

Houkes, Wybo, and Pieter E. Vermaas. 2010. *Technical Functions*. Dordrecht: Springer.

Hughes, Thomas P. 1986. "The Seamless Web: Technology, Science, Etcetera, Etcetera." *Social Studies of Science*, 16: 281–292.

ITER Organization. 2019. https://iter.org.

Janssens, Koen, ed. 2013. *Modern Methods for Analysing Archaeological and Historical Glass*. 2 vols. New York: Wiley.

Kant, Vivek, and Eric Kerr. 2019. "Taking Stock of Engineering Epistemology: Multidisciplinary Perspectives." *Philosophy & Technology*, 32, 685–726. doi:10.1007/s13347-018-0331-5

Kroes, Peter. 1992. "On the Role of Design in Engineering Theories." In *Technological Development and Science in the Industrial Age*, edited by Peter Kroes and Martijn Bakker, 69–98. Dordrecht: Kluwer.

Latour, Bruno. 1987. *Science in Action*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.

Leonard, Anthony W. 2014. "Edge-Localized-Modes in Tokamaks." *Physics of Plasmas*, 21: 090501.

Lyotard, Jean-François. 1984. *The Postmodern Condition: A Report on Knowledge*. Minneapolis: University of Minnesota Press.

Meijers, Anthonie, and Peter Kroes. 2013. "Extending the Scope of the Theory of Knowledge." In *Norms in Technology*, edited by Marc J. de Vries, Sven Ove Hansson, and Anthonie W. M. Meijers, 15–34. Dordrecht: Springer.

Mirowski, Paul, and Esther-Mirjam Sent. 2008. "The Commercialization of Science and the Response of STS." In *The Handbook of Science and Technology Studies*. 3rd ed., edited by Edward J. Hackett, Olga Amsterdamska, Michael Lynch, and Judy Wacjman, 635–689. Cambridge, MA: The MIT Press.

Nightingale, Paul. 2009. "The Nature of Technological Knowledge." In *Handbook of Philosophy of Technology and Engineering Sciences*, edited by Anthonie W. M. Meijers, 351–374. Amsterdam: Elsevier.

Niiniluoto, Ilkka. 1993. "The Aim and Structure of Applied Research." *Erkenntnis* 38: 1–21.

Nörstrom, Per. 2011. "Technological Know-How from Rules of Thumb." *Techné* 15: 96–109.

Peters, Fabian, Karsten Schwarz, and Matthias Epple. 2000. "The Structure of Bone Studied with Synchrotron X-Ray Diffraction, X-Ray Absorption Spectroscopy, and Thermal Analysis." *Thermochimica Acta* 361: 131–138.

Radder, Hans. 2009. "Science, Technology, and the Science-Technology Relationship." In *Handbook of Philosophy of Technology and Engineering Sciences*, edited by Anthonie W. M. Meijers, 65–91. Amsterdam: Elsevier.

Royal Academy of Engineering. 2018. "This Is Engineering." https://www.thisisengineering.org.uk/

Rudner, Richard. 1953. "The Scientist qua Scientist Makes Value Judgments." *Philosophy of Science* 20: 1–6.

Simon, Herbert A. 1981. *The Sciences of the Artificial*. 2nd ed. Cambridge, MA: MIT Press.

Skolimowski, H. 1972. "The Structure of Thinking in Technology." In *Philosophy and Technology*, edited by Carl Mitcham and Robert Mackey, 42–49. New York: Free Press.

Stanley, Jason, and Timothy Williamson. 2001. "Knowing How." *The Journal of Philosophy* 98: 411–444.

Staudenmaier, John. 1985. *Technology's Storytellers*. Cambridge, MA: The MIT Press.

Tomorrows Engineers. 2019. https://www.tomorrowsengineers.org.uk/students/what-is-engineering/

Van de Poel, Ibo. 2009. "Values in Engineering Design." In *Handbook of Philosophy of Technology and Engineering Sciences*, edited by Anthonie W.M. Meijers, 973–1006. Amsterdam: Elsevier.

Van Eck, Dingmar. 2015. "Mechanistic Explanation in Engineering Science." *European Journal for Philosophy of Science* 5: 349–375.

Vincenti, Walter. 1990. *What Engineers Know and How They Know It*. Baltimore: Johns Hopkins.

Von Wright, Georg H. 1963. *Norm and Action*. London: Routledge & Kegan Paul.

Wilholt, Torsten 2006. "Design Rules: Industrial Research and Epistemic Merit." *Philosophy of Science* 73: 66–89.

Zwart, Sjoerd D. 2019. "Prescriptive Engineering Knowledge." Forthcoming. In *Routledge Handbook for the Philosophy of Engineering*, edited by Neelke Doorn and Diane Michelfelder. London: Routledge.

CHAPTER 8

THE EPISTEMIC ROLE OF TECHNICAL FUNCTIONS

BETH PRESTON

1. Introduction

In the philosophy of technology literature, technical functions of artifacts are routinely distinguished from their social functions and from the biological functions of living things. Roughly, technical functions are utilitarian functions of technology, usually centered on physical transformations, such as cutting up vegetables or transporting people from one place to another. Social functions, in contrast, are functions of technology centered on managing social relationships or status, such as serving as currency or identifying someone as a police officer. It is common for technologies to have both kinds of function simultaneously. For example, in feudal Japan samurai wore two swords, one long and one short, in a combination called *daishō*. These swords identified them as samurai while still serving as very efficient weapons. Finally, biological functions are usually understood as the evolved performances of the bodily parts of plants, animals, and other living things, such as wings for flying or leaves for photosynthesizing. In using these distinctions, then, we assume that technical functions are a kind or category of function, ontologically distinct from other kinds such as social or biological functions. This chapter asks about our classification of functions into kinds. How do we carry out this classification? How *should* we carry it out? And if we carry it out the way we should carry it out, what is the status of technical functions as a kind?

In Section Two, I review some cases that reveal a general problem for the classification of functions—what I call the continuum problem. In Section Three, I argue that this is a special case of a longstanding debate about classification and natural kinds in philosophy of science, and I recommend looking to the current state of this debate for a solution to the continuum problem in the case of function kinds. I then explain, in Section Four, that this solution calls for us to consider classification as methodology—that is, as an epistemological project, not a purely ontological one. I return to one of the original problem cases from Section Two to demonstrate how considering classification as methodology can resolve the continuum problem. Finally, in Section Five I discuss technical functions as a kind of function distinct from social functions and biological functions. I argue that the methodological disadvantages of classifying functions into these kinds outweigh the advantages.

2. The Continuum Problem

In this section I outline two cases where classification of functions into kinds has been questioned—in both cases because of an underlying difficulty I call the continuum problem.

2.1 Preston on Proper Function and System Function

In my first foray into the function literature (Preston 1998), I proposed a pluralist theory that distinguished two main kinds of function—proper function and system function. Although my main interest was in a theory of function for artifacts, I based my argument for this classification on a parallel distinction in biology between adaptation and exaptation, proposed by Stephen Jay Gould and Elizabeth Vrba (1982). On their view, an adaptation is a biological feature that has been selected for its current role, and its operation is its function. An exaptation, however, is a feature that does have an effect—a current role—but has not been selected for it. It may have been selected for some other role, or for no role, but has been pressed into service under current conditions. Once pressed into service, the effect it has, may, of course, come under selection pressure, thus eventuating in an adaptation down the road. For example, feathers originally were selected for thermoregulation, and then exapted for gliding through the air, resulting eventually in some kinds of feathers being selected for flight.

I adopted the term "proper function" from Ruth Millikan (1984), who intended it to designate the selected performances of adaptations, and to cover not only the performances of biological traits but of culturally selected artifacts as well. I added a complementary term—"system function"—to designate the unselected performances of exaptations, relying for theoretical backing on Robert Cummins' (1975) analysis of function as current causal role in a system. He, too, intended his view to cover both biological organisms and artifacts. For example, chairs have the proper function of supporting seated humans—that is what they are designed and made for—but they often have the system function of step stools—for instance, when you stand on a chair to reach items on a high shelf. Note that in this particular case, the system function persists without coming under selection. Chairs are still made exclusively for sitting on, not standing on. I called this an "ongoing system function" to indicate that some system functions never evolve into proper functions, although others do. On my pluralist theory, then, there are two kinds of functions. Proper functions are established by a history of selection and reproduction for the performance that constitutes their function. System functions are established by current capacity to perform a role in a given system, regardless of history or selection pressure.

Daniel Dennett (1998) published a forceful rejoinder to my view, pointing out what he regarded as a fatal flaw. He couched his criticism in terms of adaptations and exaptations, but it translates readily to proper functions and system functions. Dennett argues that this is a distinction without a difference. Every proper function must start out as a system function, because the performance that will eventually constitute the proper function must be available for selection to act upon. For example, if feathers are eventually to have the proper function of flight, some of them must already have some serendipitous aptness for aiding flight on which natural selection can act. But, Dennett argues, this just means there is no bright-line distinction to be had between proper and system functions. System functions are typically on their way to being proper functions, so there is an unbroken continuum, not a joint in nature that would legitimate classification into two kinds of function. We may call this the continuum problem.

But what about ongoing system functions, such as the common use of chairs as step stools in which no new proper function ever emerges? Here again, Dennett argues, there is a continuum, although of a somewhat different kind. Suppose we agree that the use of anything for a purpose constitutes a system function. Then not only every artifact we use for a purpose that is not its proper function, but also every naturally occurring object we use for any purpose whatsoever automatically has that use as a system function. But this means that eggs or antelope haunches, when eaten, have the system function of providing nutrition for humans; the air, when breathed, has the system function of providing oxygen; and every stone stepped on is a system-functional stepping stone. Dennett regards this as a *reductio ad absurdum* of my view. And indeed, it is a persistent criticism of Cummins' original view. Only the rather diaphanous requirement of a containing system limits this implausible proliferation of system functions, which is grounded in the continuum between artifactual and naturally occurring things of which we make use.

2.2 Sperber on Biological Function and Cultural Function

One more quick example will help make the point about the ubiquitousness of the continuum problem for proposed function classification schemes. Dan Sperber (2007) has argued that the distinction between artifacts and naturally occurring objects is not a well-grounded classification because there is a continuum between nature and culture. The centerpiece of his argument is a discussion of biological and cultural proper functions. He begins with the observation that biological artifacts—and here he means domesticated plants and animals, primarily—have both types of functions. If there were a joint in reality between nature and culture, you would expect these biological and cultural functions to be distinct, but they actually coincide. Specifically, Sperber argues, domesticated organisms carry out their cultural functions in virtue of carrying out their biological functions. For example, wheat carries out its cultural function as food by carrying out its biological function of reproducing through producing seeds. Moreover, the reverse also holds—wheat carries out its biological function of reproducing by providing food for humans. Although we eat some of the seeds, we more than make up for this by protecting and planting the rest. And natural selection is quick to take advantage of this opportunity by modifying wheat seeds to make them even more attractive to us. This coincidence of biological and cultural functions in domesticates, Sperber argues, shows that far from being the locus of a legitimate classificatory divide between nature and culture, the realm of domestication is the locus of their imperceptible merger. And this means the imperceptible merger of biological and cultural functions as well. So if by "technical function" we mean a kind of function arising in human culture as opposed to in biology, we have no way to draw the line between these two kinds of function because of the demonstrable continuum between culture and biology.

3. The Classification Problem

Arguments based on the continuum problem are grounded in the assumption that distinguishing kinds is a thoroughly ontological operation. What we are trying to do, as it is often said, is to carve the world at its joints. On this assumption, a continuum is a problem because it demonstrates that there is no joint on which to focus our carving efforts. A continuum of cases is thus incompatible with the existence of legitimate kinds. We can, of course, carve the continuum up any way we like, but this will at best be a pragmatic or epistemic operation, not a properly ontological one.

However, this seemingly innocuous assumption is not warranted. It ignores developments in the understanding of natural kinds, especially prominent in philosophy of biology over the last few decades. A natural kind, as traditionally understood, was a group of things that belong together independently of any human interest or purpose in grouping them that way. Kinds were understood as defined by a fixed and immutable essence, shared by all the members of the kind, that distinguished them from members of other kinds. Biological species were traditionally taken to be a paradigm case of natural kinds, so understood. There are a number of reasons to doubt this essentialist view of species (Ereshefsky 2016), but one of the main reasons is grounded in the continuum problem (Hull 1965). On the essentialist view, there can be no continuum of cases between species. But if—as Darwin argued and most of us now believe—new species evolve by incremental variation out of existing species, then there *is* a continuum of cases between species. So either species are not natural kinds, as traditionally assumed; or they are natural kinds, but we must understand biological natural kinds in a non-essentialist way that accommodates continua.

Both options have proponents. Michael Ghiselin (1974), for example, argues that species are not natural kinds with individuals as members, but rather individuals with organisms as parts. But the second option has proven to be the more popular one, with a number of different views on offer. Paul Griffiths (1999), for example, proposes a kind of relational essentialism, which takes relations between organisms and other organisms rather than intrinsic properties to define natural kinds. Alternatively, Richard Boyd (1999) proposes that natural kinds are defined by homeostatic property clusters—relatively stable groupings of properties held together by homeostatic mechanisms of various sorts. In the case of species, for example, a common natural selection regime would be one of the homeostatic mechanisms holding the properties of a species together over time.

This discussion in philosophy of biology has spilled over into a general discussion about natural kinds in science. This is because kinds in many other areas of science also suffer from the continuum problem and other difficulties that make traditional essentialism about natural kinds untenable. Muhammad Khalidi (2013, chapter 5) details cases in the chemical, biological, physiological, and social sciences where widely accepted kinds are "fuzzy," or have graded membership, for instance. In response, he advocates an account of natural kinds that incorporates the influence of human interests and epistemic concerns, while still insisting that these interests and concerns are constrained by features of the world identifiable by science. Khalidi's account joins a growing list of non-essentialist accounts of natural kinds, according to which kinds are real, but their reality does not require that they be defined in total isolation from human beings, their activities, interests, epistemic projects, pragmatic concerns, and so on. As John Dupré puts it:

> My thesis is that there are countless legitimate, objectively grounded ways of classifying objects in the world. And these may often cross-classify one another in indefinitely complex ways. Thus while I do not deny that there are, in a sense, natural kinds, I wish to fit them into a metaphysics of radical ontological pluralism, what I have referred to as 'promiscuous realism.'
>
> (Dupré 1993, see also 1981)

This development in philosophy of science has led to an epistemological turn in the understanding of natural kinds. As Thomas Reydon (2014, 133) explains, we can understand natural kinds as epistemically successful categories, anchored in features of the world but dependent for their specifics on the epistemic context of science and what scientific investigation requires in terms of classification. Reydon goes on to apply this point to artifact kinds. Although there is not a science of artifacts per se, he looks to future developments in philosophy of technology and recommends that they draw on frameworks such as that provided by Boyd's homeostatic property cluster (HPC) to delineate artifact kinds. This approach, and the epistemological turn on which it is based, are particularly suited to debates about artifact kinds which—in contrast to biological kinds, for instance—have traditionally been understood as subject to human interests and projects.

In a similar vein, this same turn has been characterized by Catherine Kendig, John Dupré, and others as a turn to practice (Kendig 2016). Their point is that the turn is even more methodological than epistemological. It is not just about how scientists know the world, but how their activities shape it by classifying the subject matters of their investigations. We would do better to think in terms of activities of kinding or classifying, rather than in terms of carving nature at its joints. Kinds are not just discovered, they are made. And we have very little understanding currently of how that is done. So we can profitably focus on methodological questions about scientific activities rather than on strictly metaphysical questions about the world's contribution to the effort. Which is not to deny the world's contribution—just to place it in an appropriate perspective in light of the recent developments in philosophy of science with regard to the sticky issue of natural kinds.

I think the best way to characterize this turn is as a methodological turn. Methodology incorporates the practice concern because a methodology typically specifies the methods, principles, and processes to be deployed in a specific investigation or discipline. Similarly, methodology incorporates the epistemological concern, because the methods specified are supposed to be appropriate to the epistemological situation and aims of the investigation or discipline. I will therefore refer to the turn as the methodological turn in what follows.

4. The Methodological Turn and Function Kinds

The methodological turn does resolve one issue with regard to the continuum problem and arguments based on it. The focus on methodology embodies the assumption that traditional essentialist accounts of natural kinds must give way to non-essentialist accounts. In general, non-essentialist accounts hold that natural kinds—and *a fortiori*, non-natural kinds—are picked out in part in terms of human interests and activities, not solely in terms of objectively identifiable features of the world. The first thing to notice about this view is that it renders arguments against specific classification schemes based on the continuum problem unsound. Such arguments assume that identifying a continuum between categories *ipso facto* shows that the categories are illegitimate. Daniel Dennett, for instance, sought to invalidate my distinction between system functions and proper functions simply by pointing out the continuum between them. But we can now see that Dennett's argument—and all similar arguments—are over-hasty in their conclusions. Continua do indicate that the world does not by itself determine where classificatory lines must be drawn. But continua are not featureless; and the features they exhibit show us where lines *can* be drawn, which allows us to ground perfectly legitimate classification schemes. Biologists and philosophers of biology still have many legitimate uses for species, after all, in spite of the evolutionary continua that connect them.

More importantly for our purposes, the methodological turn prompts us to articulate the complementary epistemological or practical concerns that motivate drawing the lines where we want to draw them. Non-essentialist classification is not an everything-is-permitted enterprise. On the one side it must be grounded in features of the world, but on the other side—the human side—it must be grounded in productive, legitimate epistemic practices and viable knowledge projects. So for any classification scheme we propose, we should be able to say not only what features of the world it rests on, but how and why classifications grounded in those features help us move forward with the investigations we have in hand. We should always be able to answer questions such as: Why do we need this classification scheme? How does it help us advance our knowledge? What new or significant questions does it allow us to formulate? What approaches to answering them does it render more visible, or more possible?

To illustrate what I mean, I will give a brief analysis of my distinction between proper and system functions along these lines. First, let's look at the features of the world that underwrite the distinction. As explained above, proper functions are established by a history of natural or cultural selection that ensures reproduction for the performance constituting the function. In contrast, system functions are established by a current causal role in a natural or cultural system. So an identifiable history of reproduction or an identifiable causal role in a system are the requisite features of the world. But what do we need this classificatory distinction for? What does it do for us, epistemically?

First, it helps us analyze how functions are established and changed (Preston 1998, 2000, 2013). This is especially important for artifact function, where the creation of new functions and change of function is much more common and more rapid than it is in the biological realm. Consider new functions first. As Dennett rightly pointed out, every proper function begins with a system function—that is, with a performance that is not yet a proper functional performance because it has not yet acquired the history of selection and reproduction that defines proper functions. For example, the first working prototype of an adding machine performed additions, but doing so was not its proper function. Nevertheless, we want some way of talking about this working prototype as functional—indeed, as performing the same function as its later, proper functional descendants. The concept of system function gives us a way of doing this. Now consider change of function. Sometimes an original proper function is lost as another one is acquired. This is arguably the case where things take on a ritual or ceremonial function—trophy cups, for instance. More commonly, the new proper function is layered over the original one—for example, some diamond rings are made specifically as engagement rings, with the additional proper function of indicating the prospective marital status of the wearer. In such cases, the new or additional performances are not proper functions until the thing starts being reproduced for them, so we need a concept like system function for the same reason we need it in the case of a novel proper function. Perhaps the most significant changes are where the new use does not become a new proper function, but simply persists as a common off-label use, so to speak. We have already mentioned the frequent use of chairs as step stools. And there are also many cases where the new use is a one-off occurrence or idiosyncratic. Children may use chairs to build a playhouse, an indigent graduate student may use one as a table, inventive home decorators may use one as a plant stand, and so on. These are all cases where we want to talk about function, because the behavior and the effect is the same. You may succeed just as easily in retrieving your little-used punchbowl from the top shelf of your cabinet by standing on a chair as by standing on a step stool. But without the concept of system function this parallel is glossed over. Finally, we not infrequently use naturally occurring objects to serve our purposes—laying a path with fieldstones rather than paving bricks, for instance. Here again, without a concept like system function in our repertoire, the parallels in the behavior and the effects are not easy to see, let alone articulate.

Second, the distinction between proper function and system function helps us analyze the relationship between individual and society (Preston 2000, 2013). On the one hand, individuals are constrained by their society in a myriad of ways, many of them extremely subtle, as Michel Foucault, for instance, has been forceful in pointing out. On the other hand, as Foucault (1982) also notes, the exercise of power in society presupposes free subjects, who can resist the exercise of power. Foucault consequently prefers to speak of power governing subjects—steering them, that is, rather than determining their behavior. And the subject's ability to resist is at bottom the ability to govern the actions of others in turn. This description of the relationship between individual and society is very abstract. So even if basically correct, it really does not tell us much about the way this governance relationship is managed by either individuals or societies. One way of investigating this is to start by noticing that human activity is pervasively mediated by material culture. Thus power relations pervade our use of artifacts, because the relationship between the individual and the social order is articulated already in this material dimension. The distinction between system function and proper function is an important analytical tool for understanding this articulation and analyzing the embedded power relations. Moreover, it supports Foucault's conception of power as involving the permanent possibility of resistance on the part of the individuals governed by it.

The proper functions of artifacts are normative in the sense that they specify what the artifact is supposed to do, in a non-moral sense of "supposed to." But this means they are also normative in the sense that they specify what users are supposed to do with the artifacts. Tableknives are for cutting up food on your plate while you are eating. This is what you are supposed to do with them. Of course, you can do all sorts of other things with them, such as use them to remove screws, pry open cocoa tins, apply paint to a canvas, and so on. These are system functions, in my classification scheme, and they are not normative in either of the senses that proper functions are.

So one way in which systems of social order are imposed and enforced is through general insistence on using artifacts for their proper functions. This ensures norms of behavior that span cultures and persist across generations. As Foucault might put it, the proper functions of artifacts mediate the government of the actions of individuals, and therefore mediate the establishment and maintenance of the power relations in society. But this does not prevent individuals from resisting that exercise of power by using artifacts in system functional ways. Indeed, inventive uses of artifacts are widely admired and regarded as exhibiting the creativity of the user. Most system functions are relatively innocuous as resistance to power goes, of course, but we only have to remember the weaponizing of airplanes and motor vehicles by terrorists or the wearing of men's clothes by women to understand that system functions are sometimes far from innocuous. In short, because proper function is a normative, collectivity-centered concept, it is linked to the established social order; whereas system function is a non-normative, agent-centered concept, linked to the independence of the individual. Thus the distinction between proper function and system function is an indispensable tool for investigating the relationship between society and individual and analyzing the exercises of power that pervade it.

5. Technical Functions as a Function Kind

With all this in mind, let us return to the question of technical functions as a function kind. They are commonly distinguished from biological functions on the one side, and from social functions on the other. We will look at these pairwise classifications in turn. As a textual basis for our discussion, we will rely on the introduction to Vermaas, et al.'s (2011) recent book, *A philosophy of technology: From technical artefacts to sociotechnical systems*. They do a very good job of describing the most common reasons in favor of this classification of functions. They couch their account in terms of a classification of objects into technical, social and biological kinds, but this classification of objects generates a parallel classification of functions. The differences they outline between these kinds of function can stand alone as a separate classification of functions.

5.1 Technical Functions and Biological Functions

According to Vermaas et al., the major difference between natural (biological and mechanical) functions, on the one hand, and artifact (technical and social) functions, on the other hand, is that natural functions arise out of the operation of the causal laws of nature, and so are endemic to biological organisms and natural physical mechanisms. Technical and social functions, however, are imposed on artifacts by us. Vermaas et al. regard this as an updated version of Aristotle's view that everything in nature has its principle of motion and change within it, and a purpose it seeks to realize thereby. This is definitive of what Aristotle calls substances—real individual things. Artifacts, on Aristotle's view, are not substances because they have their principle of motion and change in us, not in themselves. Not only does their physical structure depend on our activity, so does their function. For example, in order to have an artifact with the function of toasting bread I not only have to form materials into the standard shape of a toaster, I also have to design it with the purpose of using it to toast bread.

Vermaas et al. go on to list some other differences between natural and artifact functions. Artifact functions lend themselves to normative claims, whereas natural functions do not. This is less true of biology than physics, of course. It would be odd to talk about a rock as being a good rock or a bad one; but we might well say that a human or a cat has good teeth or bad teeth. They also point out that we talk about both the parts of an artifact and the artifact as a whole as having functions, but in the case of biological organisms we only talk about the functions of its organs or behavioral traits, not about the organism as a whole as having any function. A tiger's teeth have a function, but the tiger itself does not. And they add that biological organisms do not have use plans for their functional parts the way human makers have use plans for artifacts. An albatross does not have a plan for using its wings the way Icarus had a plan for using his.

Now, you might well have objections to any of these ways of differentiating natural from artifact functions. For instance, it does make sense from an ecological point of view to talk about whole organisms as having functions. Tigers are apex predators, and that is a functional designation with regard to the ecosystem, for instance. But that is not what we need to focus on here. Rather, we should first note that there are indeed features of the world we can conscript to underwrite a dividing line between biological and artifact functions. This is exactly what Dupré's promiscuous realism predicts—there are lots of potential dividing lines in nature, and the chances are very good we will be able to draw an ontologically plausible line where we need it for epistemic purposes. Second, we should note that no matter where we draw our lines, we will typically find them bridged by a continuum of cases. Vermaas et al. acknowledge as much in the case of the line they wish to draw between natural and artifact functions. After listing a number of equivocal examples—functions associated with genetically engineered organisms, natural objects used for the same purposes as artifacts, etc.—they say:

> The dividing line between the natural and artificial worlds is a sliding scale; there is no clear-cut division between the two. Yet, that does not mean that there is no clear difference between paradigmatic examples of natural objects and technical artefacts. As we have seen, those differences do exist and, to sum up, those differences relate especially to the status of having a function and a use plan, and to the accompanying possibility of making normative assertions. (Vermaas et al. 2011, 11)

But focusing on the clear cases amounts to a *choice* in favor of those cases as paradigmatic and against the equivocal cases as insignificant. The world offers us a plethora of features we can use to construct our ontologies, but does not constrain us to a uniquely correct one, either in the short run or in the limit of inquiry. So in the end, we must acknowledge our own contribution to ontology, and ask for the rest of the story—the epistemic part of it, in other words. Why do we need this classification scheme? How does it help us advance our knowledge? What new or significant questions does it allow us to formulate? What approaches to answering them does it render more visible, or more possible?

Vermaas et al. offer us no explicit help with this, but the fact that they are engaged in philosophy of technology provides some clues. Maintaining a distinction between artifacts and natural objects—and therefore between artifact (technical/social) functions and natural (biological/mechanical) functions—may be epistemically important to delimiting the field of investigation for this area of research. Notice that the distinction between system and proper functions has no usefulness in this regard, because it is domain neutral. Functional objects in any domain may have system and/or proper functions. But the classification system for functions that Vermaas et al. advance is domain specific. It delimits domains of objects—and therefore epistemically relevant domains of research—in part by classifying functions into kinds. This is a general epistemic operation of domain specific function classification schemes. In addition, Vermaas et al. might also claim that their classification scheme provides the epistemological foundations for certain aspects of action theory—those concerned with the activities of design and production. Indeed, the philosophy of technology advanced by Vermaas et al. revolves around the concept of engineering, which they take to be central to understanding technical functions and the use plans that accompany them. The special characteristics of artifact functions may help articulate the special characteristics of engineering as a human pattern of activity.

But the connection with philosophy of technology and engineering also raises some troubling questions about the epistemic disadvantages of this function classification scheme. Dan Sperber ends his previously mentioned piece with a specifically epistemological worry.

> Here I have tried to cast doubt on the idea that a theoretically useful notion of artifact can be built around its usual prototypes: bracelets, jars, hammers, and other inert objects, or that it can be defined in a more systematic way . . . . There is no good reason why a naturalistic social science should treat separately, or even give pride of place to, cultural productions that are both more clearly intended for a purpose and more thoroughly designed by humans, that is, to prototypical artifacts.
>
> (Sperber 2007, 137)

Sperber's main argument for this conclusion is based on the continuum problem, especially as it concerns what he calls biological artifacts, such as domesticated plants and animals. As we have noted, this is not sufficient, because all classification schemes face such continua. But Sperber does supply an interesting epistemological consideration as well. He suggests that in focusing on paradigmatic artifacts as the basis for our theories, we are allowing ourselves to be deceived by "a doubly obsolete industrial-age revival of a Paleolithic categorization" (136). He explains, first, that in the Paleolithic, before there were any domesticates other than dogs, the vast majority of the technologies people used in their daily lives *were* the allegedly paradigmatic type of artifacts—stone tools, baskets, beads, and so on. So we may well have evolved a psychological disposition to classify things in accordance with the salience of such artifacts—a disposition we now have trouble shaking, even though the Neolithic transition to agriculture 12,000 years ago made biological artifacts (as Sperber calls domesticates) proportionally the most common type of artifact in our experience until the Industrial transition of only a couple of centuries ago. Second, Sperber argues, information technology and biotechnology are increasingly contributing to our environment artifacts that would have astonished Aristotle in their ability to act and "think" on their own, beyond any design or intention their creators may have. They more nearly resemble domesticates in this regard than the "inert" paradigmatic artifacts Sperber describes. And they, too, were unimaginable to our Paleolithic ancestors, whose lagging psychology and classification schemes we have inherited, and which are now distorting our epistemic perspective on the world. The well-being of our science and philosophy may therefore depend on our resisting the urge to draw classificatory lines in the time-honored place where nature merges with culture, and to discount this liminal zone as of little interest to our inquiries.

We may add to Sperber's general epistemic uneasiness about the social sciences a more specific worry about the state of the art in philosophy of technology. The biological artifacts he identifies are a technology—agricultural technology, to be precise. And agriculture is our subsistence technology. From it we derive not only the preponderance of the world's food, but also materials such as fiber, wood, and increasingly, fuel. Moreover, agriculture is not a technology we can get along without—at least not at present population levels. Add to this the worries—voiced early on by Rousseau (1997) in his *Discourse on the Origin of Inequality* and echoed by anthropologists, if not philosophers, since—about the role of agriculture in the rise of hierarchically organized societies that disadvantage many of their members in order to disproportionately advantage a small elite. Yet with the exception of a few people like Paul Thompson (2010, 2017) and Gary Comstock (2000), agriculture is a neglected subject in philosophy of technology. And even they tend to focus on recent developments in agricultural technology, such as factory farming and genetic engineering, rather than looking at the whole sweep of agricultural history and the diversity of forms agriculture has assumed. From this point of view it is difficult to avoid the conclusion that the epistemic choice Vermaas et al. and so many others make to focus on "clear" examples of technology like bridges and hammers risks serious epistemic distortion with regard to our technological situation.

5.2 Technical Functions and Social Functions

So far we have been considering a distinction between technical and social functions on one side—the artifact side—and biological functions on the nature side. Now we need to consider the internal distinction on the artifact side between technical and social functions. Both of these kinds of functions depend on us, but in different ways. As Vermaas et al. put it:

> Technical artifacts fulfill their function by virtue of their physical properties whilst social objects depend for their function upon their social/collective acceptation. (Vermaas et al., 12)

For example, they say, an airplane fulfills its function of transporting people and cargo in virtue of its physical structure and the materials of which it is made. What anyone thinks of that function is irrelevant to whether the plane is able to fulfill it. Money, however, can fulfill its function of serving as legal tender in transactions between buyers and sellers only if it is generally accepted that it has that function. Here, the materials and physical structure of the money are not relevant, which is why legal tender can take so many different forms. So the function of the airplane depends on us only in the sense that we build airplanes. In the case of money, we not only print bank notes, but we sustain their function by continuing to accept that they are legal tender for purchases.

As with the distinction between artifact and natural functions, Vermaas et al. acknowledge that there is a continuum problem. But it has a slightly different cast in this case. They use the example of traffic lights to explain that although there is a necessary physical structure underwriting their functioning, it is also necessary that there be duly enacted and promulgated traffic laws adhered to by the citizenry for the traffic lights to actually function as intended—that is, as regulators of traffic. So in this case the continuum is actually internal to the artifact, which Vermaas et al. call a sociotechnical system. Although they do not mention it, the further problem is that all artifacts are sociotechnical systems in this sense. Take the airplane—it functions as transport not just because it has a certain physical structure, but because people accept it as a type of transport, and sign up as passengers or crew members. But in any case, Vermaas et al. argue, the distinction between technical and social functions still holds, because it is possible to pick out the social and technical aspects of an artifact, and analyze them separately. And they are certainly right that there are distinguishable features of the world that allow you to do this. If you focus on the features of the world that concern the physical structure of an artifact, you can reasonably claim to be analyzing its technical function; whereas if you focus on the features of the world that concern the psychology and practices of its users, you can claim to be analyzing its social function.

However, there are other possible classification schemes for artifact functions that focus on alternative features of the world. Michael Schiffer (1992, 9–12), for example, distinguishes technofunctions, sociofunctions and ideofunctions. The technofunction is the utilitarian function—a chair, for instance, has the technofunction of supporting seated humans. The sociofunction manifests social facts—an expensive chair by a well-known designer manifests the socio-economic status of the owner, for instance. And the ideofunction involves the symbolizing of abstract values or beliefs—a throne, for instance, is a special kind of chair that symbolizes ruling authority. There is obviously some overlap between Schiffer's classification scheme and that of Vermaas et al., but they do not completely align. This shows that pointing to features of the world to legitimate a classification scheme is not sufficient. So we must ask: what are the methodological implications of any given scheme?

Vermaas et al. might well claim that their classification scheme channels our attention to the important task of sorting out the differential contributions of physical structure and human intention in the production and use of artifacts. It is reasonable to suspect, in fact, that this is a lot more complicated than they let on, since the term "collective acceptance" covers a multitude of complications. For example, acceptance may depend on the material make-up of the artifact. Think of the consternation among the populace when countries began using paper money rather than gold or silver coins that had some intrinsic value, for instance. And consider that even sheer refusal by a majority of the populace to countenance the change would likely not have been successful, because some members of the collective were in positions of authority over others as far as deciding what the material realization of the currency would be. Arguably, then, the distinction between social and technical functions has the methodological advantage of focusing our attention on the interaction of physical structure and human intention in the establishment and management of artifact functions.

However, here too there are reasons to worry that the methodological disadvantages of this distinction outweigh its benefits. Vermaas et al. claim that a specific kind of physical structure is a necessary condition of a technical function, whereas a particular physical structure is not necessary to social functions. Rather, they are a matter of human beliefs and intentions, constituting a collective acceptance as to the function. Here we are in uncomfortably Cartesian territory. The causal laws of nature are relevant to technical functions, but not to social functions, apparently. Later on, Vermaas et al. do concede that technical functions have a dual nature, since a use plan is also a necessary condition. Although they do not say so, it is reasonable to think that social functions, too, have a dual nature, because a physical realization is necessary. But this only raises the question of why we are classifying these as different kinds of functions, if in both cases a physical structure and various mental elements—use plans, intentions, decisions, beliefs, or whatever—are necessary conditions. The answer seems to be that in the one case the physical structure is the dominating element, whereas in the other case the mental factors dominate. But this does little more than take the edge off the Cartesian flavor of the distinction. There is still a subterranean insistence that the mental and the physical are somehow very different—like oil and water, they just do not mix even when they are both present in the same object.

The problem here is not that there is no difference between materials and structures, on the one hand, and human decisions, intentions, and the like, on the other hand. There are also lots of differences between various materials, and various ways in which humans intend things. The problem is that classifying functions into technical and social functions ontologizes the relevant differences and so gives them epistemically distorted weight in the formulation of questions for investigation and in our resulting analyses. The worry is that these distortions threaten to undo all the work done over the last century or so on both sides of the Atlantic by people like Maurice Merleau-Ponty and Gilbert Ryle to turn our thinking and intellectual practices in more fruitful, non-Cartesian directions.

A second, and related, methodological worry again concerns issues in the philosophy of technology—specifically, the idea that technologies are morally and politically neutral in and of themselves, and that it is how they are used that brings values into the picture. The standard illustration of this view is the well-known slogan of the gun rights movement in the United Sates—guns don't kill people, people kill people. Like most other contemporary philosophers of technology, Vermaas et al. are emphatically not in favor of this neutrality thesis, and argue against it on a number of grounds (2011, 16 ff.). But arguably, their prior classification of artifact functions into technical and social predisposes in favor of it. Ontologizing the difference between the utilitarian and the social contributions to the function of an artifact suggests that artifacts can in fact be neutral; that we can ask only what they do and how they work, and that we can separate this from any inquiry into whether what they do when they work the way they are supposed to work is good or bad. Here again, the domain-specific classification of functions suggested by Vermaas et al. threatens to counteract the efforts of everyone from Karl Marx to Michel Foucault to Bruno Latour to turn our thinking and intellectual practices in the direction of a more sophisticated understanding of the role of material culture in human existence.

6. Conclusion

I have argued for two main claims in this paper. First, I have argued that the aim of classification schemes is not only to highlight recognizable features of the world, but also to serve methodological purposes. This must be the case, because the world offers us a plethora of features, and so a number of legitimate, but different and possibly crosscutting, ways to classify things into kinds. Second, I have argued that when this understanding of classification is taken into account, there are methodological reasons to worry about the standard classification of functions into biological, technical, and social functions. There may indeed be methodological reasons in favor of this classification. But we must explicitly weigh them against countervailing reasons when proposing a classification scheme for functions.

Acknowledgment

I presented an earlier version of this chapter at a workshop entitled *Artifactual, Biological, and Social Functions: Historical and Contemporary Perspectives* at the University of Bern in June, 2017. I would like to thank the other workshop participants for their very helpful questions and comments.

References

Boyd, Richard. 1999. “Homeostasis, Species, and Higher Taxa.” In *Species: New Interdisciplinary Essays*, edited by Robert A. Wilson, 141–185. Cambridge, MA: MIT Press.

Comstock, Gary L. 2000. *Vexing Nature? On the Ethical Case Against Agricultural Biotechnology*. New York: Springer.

Cummins, Robert C. 1975. "Functional Analysis." *The Journal of Philosophy* 72 (20): 741–764.

Dennett, Daniel C. 1998. "Preston on Exaptation: Herons, Apples and Eggs." *Journal of Philosophy* 95 (11): 576–580.

Dupré, John. 1981. "Natural Kinds and Biological Taxa." *Philosophical Review* 90(1): 66–90.

Dupré, John. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Ereshefsky, Marc. 2016. "Species." *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), Edward N. Zalta (ed.). https://plato.stanford.edu/archives/sum2016/entries/species/.

Foucault, Michel. 1982. "The Subject and Power." In *Michel Foucault: Beyond Structuralism and Hermeneutics*, edited by Hubert L. Dreyfus and Paul Rabinow, 208–226. Chicago: University of Chicago Press.

Ghiselin, Michael. 1974. "A Radical Solution to the Species Problem." *Systematic Zoology* 23: 536–544.

Gould, Stephen Jay, and Elizabeth S. Vrba. 1982. "Exaptation: A Missing Term in the Science of Form." *Paleobiology* 8: 2–15.

Griffiths, Paul. 1999. "Squaring the Circle: Natural Kinds with Historical Essences." In *Species: New Interdisciplinary Studies,* edited by Robert A. Wilson, 209–228. Cambridge, MA: MIT Press.

Hull, David. 1965. "The Effect of Essentialism on Taxonomy: Two Thousand Years of Stasis." *British Journal for the Philosophy of Science* 15: 314–326 and 16: 1–18.

Kendig, Catherine, ed. 2016. *Natural Kinds and Classification in Scientific Practice*. London, UK: Routledge.

Khalidi, Muhammad Ali. 2013. *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge, UK: Cambridge University Press.

Millikan, Ruth G. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism.* Cambridge, MA: MIT Press.

Preston, Beth. 1998. "Why Is a Wing Like a Spoon? A Pluralist Theory of Function." *Journal of Philosophy* 95 (5): 215–254.

Preston, Beth. 2000. "The Functions of Things: A Philosophical Perspective on Material Culture." In *Matter, Materiality and Modern Culture*, edited by Paul Graves-Brown, 22–49. London: Routledge.

Preston, Beth. 2013. *A Philosophy of Material Culture: Action, Function, and Mind.* New York: Routledge.

Reydon, Thomas A. C. 2014. "Metaphysical and Epistemological Approaches to Developing a Theory of Artifact Kinds." In *Artifact Kinds: Ontology and the Human-Made World,* edited by Maarten Franssen, Peter Kroes, Thomas A. C. Reydon, and Pieter E. Vermaas, 125–144. Cham, CH: Springer.

Rousseau, Jean-Jacques. 1997. *The Discourses and Other Early Political Writings*, edited and translated by Victor Gourevitch, Cambridge, UK: Cambridge University Press.

Schiffer, Michael B. 1992. *Technological Perspectives on Behavioral Change*. Tucson, AZ: University of Arizona Press.

Sperber, Dan. 2007. "Seedless Grapes: Nature and Culture." In *Creations of the Mind: Theories of Artifacts and Their Representation*, edited by Eric Margolis and Stephen Laurence, 124–137. Oxford, UK: Oxford University Press.

Thompson, Paul B. 2010. *Food Biotechnology in Ethical Perspective*. 2nd ed. Dordrecht: Springer.

Thompson, Paul B. 2017. *The Spirit of the Soil: Agriculture and Environmental Ethics*. 2nd ed. New York: Routledge.

Vermaas, Pieter, Peter Kroes, Ibo van de Poel, Maarten Franssen, and Wybo Houkes. 2011. *A Philosophy of Technology: From Technical Artefacts to Sociotechnical Systems*. San Rafael, CA: Morgan & Claypool Publishers.

CHAPTER 9

REVISITING SMARTNESS IN THE SMART CITY

SAGE CAMMERS-GOODWIN

1. Introduction

This chapter critically examines the intelligence of smart city government, which often ignores experiential and practical knowledge of citizens. Smart cities are a globally attractive phenomenon. They represent the future, where the normal city is improved with the assistance of ICT (Information Communication Technology) and IoT (Internet of Things) solutions (European Commission n.d.). Across continents, cities are investing in "smartness" with the hope for far-reaching positive effects across value domains such as sustainability, safety, and efficiency. Section 2 of this chapter introduces the smart city movement.

Unfortunately, the typically envisioned smart city might not be smart for everyone. There is a risk that by simply working to digitalize pre-existing systems, longstanding inequities may never be examined, interrogated, or solved. What is considered "smart" by governments and corporations to fix might bypass the needs and wants of citizens long neglected by both institutions.[1] Section 3 explores how city decision-making may not lead to smart outcomes for all residents.

Section 4 looks backward to examine whose knowledge is historically valued. The disenfranchised, such as women and people of color have continuously been excluded from the affiliation of smartness (Cave 2020). This should be concerning when we look forward to what problems the smart city will deem smart to fix and how the smart city will source its knowledge. For what has been framed as a "problem" requiring a "solution" in modern history is not reassuring: within the past two centuries, government-backed racist, sexist, classist and ableist ideology has barred individuals from voting, practicing their own religion, speaking their own language, holding positions of power, and, in the most extreme cases, from the right to reproduce or to live.

Furthermore, technology does not guarantee the wisest, fairest, or most efficient solutions to the real problems cities face. Section 5 describes the shortcomings in relying on IoT and ICT generated intelligence. Machine learning algorithms are built by naturally biased individuals and often trained using data shaped by an unjust world. Moreover, dependence on technology does not necessarily make a city more resilient. To the contrary, a technology-dependent city is one susceptible to hacking, power outages, and software bugs. Technology may be a solution in some instances, but technology does not guarantee smartness.

Section 6 concludes this chapter with a recommendation. In order to build cities that are smart for all, it is essential to break down the biases present in traditional cities. This is an epistemic challenge because the status quo has become so normalized that it may be difficult to see how infrastructure that is harmless to some may be hostile to others.

2. The "Smart City" Movement

The European Union defines the smart city as "a place where traditional networks and services are made more efficient with the use of digital and telecommunication technologies for the benefit of its inhabitants and business" (European Commission n.d.). Such environments are expected to improve transportation, increase sustainability, accelerate city government responsiveness, advance safety, and fulfill the needs of elderly citizens (European Commission n.d.). The smart city is the opposite of an unintelligent city. It not only recognizes failures and lags, but also quickly addresses them. Values such as safety, sustainability, and efficiency dictate what knowledge domains are important, and data from these spheres drive decision making.

Other, sometimes conflicting, definitions of smart cities also exist (Galdon-Clavell 2013). Some focus more on mutual creation by varied stakeholders such as citizens and government than just information communication technology (ICT), which may allow smaller, less wealthy cities to be included in the fold. After reviewing multiple definitions of the smart city, an EU-funded smart city report came to the conclusion that a "Smart City is a city seeking to address public issues via ICT-based solutions on the basis of a multi-stakeholder, municipally based partnership" (Manville et al. 2014). Their definition includes six characteristics: smart governance, smart economy, smart mobility, smart environment, smart people, and smart living. Each characteristic is informed or improved by ICT, with the exception of "smart people" who are those capable of working with and creating ICT solutions.

These threads can be seen in multiple smart city initiatives. Spurred by the demands of hosting the 2016 summer Olympic games and global concerns of safety, Brazil invested in COR, the Rio Operations Center. COR monitors Rio de Janeiro, Brazil out of a "Decisions Center" 24 hours a day 7 days a week with the assistance of 1000 video cameras and 500 professionals that take turns in three daily shifts (Schreiner 2016, 9-10). COR allows for monitoring of city assets such as "administrative buildings, schools, hospitals, cars, bus fleets, radios, and agents in the service of the municipality" in real time (Schreiner 2016, 10). The system connects with citizens by warning in case of landslide, keeping track of demographic data, and connecting directly with populations with "heavy use of social media and apps such as Waze, Moovit, Alerta Rio, and others" (Schreiner 2016, 10).

Meanwhile, New York City was named the Best Smart City of the year at the Smart City Expo World Congress 2016 in Spain (New York City Hall Press Office 2016). The award stemmed from their effort to expand internet access, a $3 million investment in gunshot detection sensors, an $18.6 million investment in a pilot for connecting vehicles through the internet of things, an accelerator for entrepreneurs featuring 100,000 square feet (approximately 9300 square meters) of affordable space, and a set of guidelines for the deployment of smart city initiatives (New York City Hall Press Office 2016). As of 2020, the main equity clause for the NYC smart city is captured under guideline 2.8: "All data sets [. . .] should be checked for geographic, social or system-driven bias [. . .] and other quality problems. Any biasing factors should be recorded and provided with the data set and corrected where possible" (NYC Mayor's Office of the Chief Technology Officer n.d.). Such unbiasing is especially necessary since New York has an Open Data platform where anyone can access and analyze data from the city (NYC Open Data 2019).

Smart cities are an international phenomenon, each blooming in unique sociopolitical landscapes, but seemingly competing to converge upon the best solution. Smart cities can usually be divided into initiatives and outcomes. Some smart cities have yet to be built and are marketed as solutions to raise the bar for the country as a whole. Others naturally took on the title and role after decades of being tech friendly, global and wealthy. These groups are not mutually exclusive. Outcome cities can also plan seemingly grandiose initiatives that may or may not become actualized. Moreover, many smart cities are not one concrete plan but a mix of advancements that justify the label of "smart." Current initiatives include India's 2015 promise for 5 years of smart city investment, which approved 5,151 projects (Khare 2019). African countries such as Rwanda, Kenya, South Africa, and Nigeria are also investing in smart visions (Giles 2018). China hosts a mix of investment-leaning and outcome-leaning cities. The country is hoping to build more cities to solve poverty by luring people out of rural areas (Manville et al. 2014, 18). Japan and Korea are also both investing and basking in smartness. Top outcome and ongoing investment smart cities in Europe include Amsterdam, Helsinki, Barcelona, Hamburg, and Oulu (Manville et al. 2014, 68). The United States includes project outcomes such as electronic public transportations in several smaller cities and bigger tech company friendly cities such as San Francisco and New York. This is a non-exhaustive list. What these cities do have in common, however, is a sense that increased data collection and integration ICT will lead to general city improvements. Who exactly these improvements are for, however, is not always clear.

3. "Smart" for Who?

New smart city initiatives seem to grant opportunities to address urban problems with increased technology, yet the majority of smart cities that have already begun to reach the outcome stage had a head start, at least in terms of technology adoption.[2] In these cities average citizens already benefitted from ICT such as internet and smart phones. Conversely, cities planning to become smart cities where citizens lack access to ICT technology are skipping a step. An element of privilege thus predetermines which cities can be labeled as smart as well as the individuals to whom the smart city caters. Investing in IoT solutions and hosting central data is expensive, and without embedded ICT infrastructure a city is unlikely to fit the current conception of smart even if it meets most of the value requirements. One could imagine a safe, sustainable, and resilient village having a very high quality of life, yet, due to a lack of modern IoT solutions, failing to reach today's standards of "smartness."

It should come as little surprise, then, that the term smart city was marketed by IBM to describe an ideal metropolis connected through ICT (Rosatia and Contia 2016, 969). While corporations used the term "smart" to encourage adoption of their technology, governments found a way to adopt corporate technology to address problems they deem "smart" to fix, such as those listed by the European Commission. The goals of the smart city are unapologetically utopian to the extent that initiatives stemming from smart city planning often overlook those who may not fit into the utopian ideal. The tech-phobic, the racially profiled, the pedestrian who cannot afford a smart phone, the citizens with disabilities are at best seen as "edge cases" in a system that otherwise "works."[3]

By automating the city to further benefit the average citizen (or in some locales, the most *valued* citizens), outliers may be left out of the smart city narrative. Improving the *traditional* infrastructure of the city is not enough to make the city ubiquitously smarter, because the original foundation may have already failed to provide necessary services to all groups. Examples of government-led initiatives to improve cities making living conditions worse for subsections of the population have been abundant even before the introduction of IoT infrastructure.

For example, Los Angeles is infamous for redlining, a practice dating back to the 1930's when the US federal government mapped racialized neighborhoods and ranked them for investment risk, enabling white home buyers to access favorable mortgages in white-only neighborhoods and restricting the options of people of color to renting or predatory loans. Redlining also made it inevitable that LA's growing Black and Latino populations would be subjected to live in areas made hazardous by the oil extraction that made LA rich in the first place (Cumming 2018). An often cited example in the politics of artifacts is the New York parkway bridges made too low for intercity busses to clear, therefore making a bus route from New York City neighborhoods to the Long Island beaches impossible (Winner 1980; see also Joerges 1999). Whether through intention or ignorance, the height of the bridges made the public beach largely inaccessible to many low income urban families. Further examples of purposeful exclusion or ignorance of urban citizen needs are still commonplace today.

"Hostile architecture" describes city infrastructure that constricts personal freedom in public space (Rosenberger 2020). Such structures include city benches with armrests that make it impossible to lie down, decorative boulders placed in locations common for homeless encampments, and metal pins to prevent skateboarders from grinding (Paulas 2019). It has been fairly pointed out that consequentialist reasoning might make such architecture defensible in some instances (de Fine Licht 2017). Nonetheless, it makes little sense to constrict the flexibility of multiuse structures if proper infrastructure for shelters, skateboarding, and other community based engagement are unavailable. Redlining and hostile architecture are both government funded initiatives. These were supposedly "smart" decisions.

If technology-dense cities want to be smarter, the first question that should be asked is "smart for whom?" It should be noted that smart *citizenship* is often either 1) absent from the smart city narrative; 2) included as a superficial element of the design process to make the inevitable outcome more palatable; 3) marketed as a side feature with limited funding in comparison to the larger government-led initiatives; or 4) integral to the smart city but only accessible to certain populations. Toronto's Sidewalk Labs, a subsidiary of Alphabet, Google's parent company, is a stark example of a smart city initiative that failed to effectively engage with citizens. Many residents were either against the project or had privacy concerns that they felt went unaddressed, even though from the start of the project Sidewalk Labs consulted with residents. Sidewalk Labs however was never truly transparent in their design plans. In early 2019, *Toronto Star* reporter, Marco Chown Oved revealed from leaked documents that the eventual plans for the project encompassed 350 square acres, as opposed to the 12 square acres that were being communicated to the community (Oved 2019). In March 2020 Sidewalk Labs decided to discontinue the waterfront project, but the project CEO cited financial concerns due to the global pandemic rather than activist demands (Doctoroff 2020). The fact that the project got so far despite wide-scale protest should warn how powerless the minority might be against the whims of the smart city.

When civic values clash, the smart city with its added power and control is primed to win. In 2019, Hong Kong activists began fighting against a bill that would allow law breakers to be extradited to mainland China (Purbrick 2019). From wearing face masks (which were banned in protest in 2019) to tearing down smart lamp posts, citizens tried to avoid possible facial recognition in a smart city that opposes the right to protest (Yang 2020). In this context, the harms to protesters identified by "smart" facial recognition technology are severe—a Hong Kong law left from colonial anti-communist British rule defines a riot, punishable by maximum 10 years jail time, as any group of three or more that disturbs the peace (Purbrick 2019). After many of the lead activists were imprisoned, protestors linked together to form "smart mobs," groups linked through livestream, WhatsApp, Telegram, and other platforms to form a quick-moving, leaderless force (Ting 2020).

According to the government, the majority peaceful protestors were the ones inciting violence, not the police force armed with rubber bullets and pepper spray. In such cases, the values the city considers "smart" primarily apply only to those activities that protect the status quo. Community safety understandably trumps individual freedom. Unfortunately, this may extend to the safety of the existing *infrastructure*, buildings, laws and statues, trumping the well-being and autonomy of residents. The pattern of traditional government and corporate values being held in higher regard than the values of the local community, may reduce opportunities for growth that increase well-being for segments of the population.

4. Systemic Exclusion from "Smartness"

Given the documented legacy of exclusion and discrimination in modern theories of and metrics for intelligence, the issue of who in the city gets to be or define what is "smart" cannot be addressed without a racial and feminist critique (Cave 2020). This question persists not only between individuals, but as well on a societal level. During the modern colonial era, dating from the early 16th through 20th century, functioning indigenous communities were labeled as savage and uneducated. This practice allowed colonists to validate overthrowing local government, introducing Christianity, looting national treasures, and mandating their own language, clothing, and constitution as law. African people were described as scientifically inferior to make the practice of slavery less heinous. Slaves were forbidden to speak their own languages (to prevent uprisings), but also legally barred from learning to read their captor's language. Slaves were forbidden to practice their own religion, but also excluded from positions of authority in the church.

These laws are not ancient. From 1810 until 1917 the US federal government subsidized boarding schools to "civilize" Native American children, separating them from their families, languages, traditions, and culture (Adams 1997). The segregation of Blacks from White schools was legal in the United States until the Supreme Court unanimously decided in the mid-1950s that such exclusionary practices were unconstitutional (*Brown v. Board of Education* 1953). Legalized supremacy of the minority white population did not end in South Africa, a former Dutch then British colony, until 1990.

The irony of who gets to be smart is apparent. Much of the wealth gained from colonization was due to the genius of indigenous populations. Native Peruvians cultivated land and invented the resilient potato as well as transferrable farming techniques that helped lift Western Europe from famine (McNeill 1999). The tomato, potato, corn, common bean, pepper, and tobacco plant, did not just naturally gift themselves in their final form to the Americas, but were bred by indigenous populations (Rasmussen et al. 2020). These techniques and their byproducts were later taught and spread to the rest of the world. The core of much international cuisine is due to the science of people who were labeled savage in colonial propaganda. Meanwhile, Black Americans were the wealth makers, cultivating land on plantations, cooking, running the home, care taking for children, providing medicine, and midwifing, tasks few would entrust to someone truly incapable.

Globally, women have been excluded from financial, governmental, and academic institutions based on the misogynistic notion that women were incapable of intelligent decision making. Most women did not gain the right to vote until World War 1 (Russia 1917, Canada 1918, Netherlands 1919, United States 1920), likely because women moved into factory work to help with the war effort, effectively dispelling the fear that women were too inept for the public arena. Still, France did not extend this right to women until 1944, Greece 1952, and Switzerland 1971. The problem here is twofold: that women were seen as incapable of being able to perform in traditional roles of intelligence, but also that roles traditionally ascribed to women are not considered smart. Alison Adam argues in *Artificial Knowing* that the requirements for artificial intelligence have been built around traditionally masculine notions of intelligence. Playing chess was the gold standard in AI for a while as opposed to tasks like managing a household (Adam 1998). Researchers are now working to improve AI performance at tasks such as care taking, therapy, and communication, but these abilities are most often associated in the literature with "humanness" as opposed to smartness.

Efforts to enforce the hegemony of the "smart" classes of society extended to 20th century eugenics. Eugenics targeted the disabled, LGBTQ individuals, people of color, the poor, and women. The mentally ill, deaf, blind, epileptic, and incarcerated were all legally targeted groups for forced sterilization in the United States, where from 1907 through to the 1970s, over 60,000 citizens were sterilized in an effort to better society (Lombardo 2010). Many of the cases were women, often poor, seen as "feebleminded" or promiscuous. What grew in the US also spread into Sweden, where between 1935 and 1975 approximately 63,000 people were forcibly, coercively, or willingly sterilized, over 90% of which were women (Government of Sweden 1992, 33).[4] In Germany between 1934 and 1945 360,000 individuals including the mentally ill, disabled, and children with African Ancestry were sterilized (Weindling 1989, 533; Kestling 1998). This metric does not include the 200,000–300,000 German children and adults killed in "euthanasia centers" before the mass murders in concentration camps began (Grodin, Miller, and Kelly 2018). In all these cases the state assumed it was smarter than groups of individuals and that the next generation of the state would be better without some people. It is important to note that these governments were not abstract bodies, but organizations constructed of individuals who materially benefitted from the undermining of others.

One might see these legacies as largely unrelated to the smart city and merely as historical mistakes. Perhaps "smart" city is just an unfortunate branding. Smart phone, smart car, smart TV—these are just naming conventions for elevated electronic systems. Perhaps it is simply comforting to associate the smart city with technologies streamlined by increased datafication and improved design. The issue is that the city is not another new singular technology. Cities have long been innovative and made decisions that were smart for some. Every continent has tours to see the ingenuity of the creators behind what are now ancient ruins. There are cities from thousands of years ago that already had operational sewer systems and democratic governments. Furthermore, if to be smart is to be connected to new multinational technology platforms, there is a danger in a city becoming another item one can just buy. One of the strongest objections to the Sidewalk Labs Waterfront Toronto project was that it was a corporate-backed project. Why does a city become smart once a tech company involves itself, or more data is collected? Would it not be smarter to reach value-driven goals without heavy reliance on surveillance or expensive technology?

Smart cities are unabashedly a marketing tactic to sell a utopian vision of a tech-company friendly city worthy of investment. While some corporate and government interests may be admirable, the history of who gets to be categorized as smart suggests the real possibility that marginalized smartness may be actively ignored, undervalued, and repressed. Epistemological literature teaches us that there are different types of knowledge. One can know how to do something, one can be aware of something as an item of knowledge (a true proposition or fact), or one can know something by direct acquaintance with it, i.e. have an experience of something (Steup and Neta 2020). In smart city logics, knowledge though experience and knowing a solution to a problem only leads to investment if it positively affects the privileged. Investments directed toward the underprivileged must be justified through studies conducted by those deemed "smart" enough to be reputable. This translates to only knowledge derived from the privileged mattering.

Standpoint epistemology offers further insight on why a small subset of knowledge is deemed universal while other knowledge is ignored. This feminist theory argues that one's position in society, which may be rooted in their gender, nationality, race, religion, etc. will shift individual knowledge. According to standpoint epistemology, there is no singular way of knowing, but a multiplicity of perspectives that grant epistemic access to reality. Given that some identities tend to have more power in society, socially privileged epistemic perspectives are granted hierarchy over others. The epistemic standpoint with the most power (due to societal bias and not superior reasoning) may be seen as essential or universal only because the dominant group holds enough power to avoid subjugation to other epistemic standpoints. Standpoint epistemologist Rebecca Kukla, among others, further argues that (1) some features of knowers, such as their social position, might grant those individuals "better, more objective knowledge" and (2) marginalized individuals might be at an epistemic advantage given that they are granted access to information and experiences impenetrable by the privileged (Kukla 2006, 81). This would suggest that those most marginalized by the city possess knowledge beyond the scope of city planners.

Regrettably, society has been molded to invalidate the lived experiential knowledge of the marginalized and deem the non-elite as less capable of intelligence. Only when AI, scholars, or government officials validate the needs of the unprivileged and conclude that fixing the concern benefits "everyone" can an investment can be rationalized. Enhancements to the status quo, however, do not need to be rationalized to the same extent in budgeting proposals. Easily recognized examples include selective investments to profitable city areas due to purported return on investment, thereby excluding poor neighborhoods from comparable enhancements. Consider the refurbishment of the Notre Dame Cathedral which raised €850 *million* through international donations after its 2019 fire and is due to finish in time for the 2024 Olympics (Cascone 2020). The magnitude of funds raised internationally and the urgency of reconstruction should be remarkable given the secular nature of the country (where wearing religious coverings is outlawed) and the fact that a perfectly reconstructed cathedral has minimal if any direct impact on the opportunities, health, and well-being of most French citizens. Proposals for issues faced by marginalized groups, in contrast, are not by default seen as "smart" solutions or investments, because smartness must improve the lives and areas seen as most important to, or as representative of, the city. This pattern reproduces the logic of the business world. Gender diversity and inclusion in hiring or board membership are claimed to matter because they are good for business (Kochan et al. 2003), not because it is good for women or people of color, who happen to make up more than 50% of the population, and moreover have been legally excluded from opportunities as explored earlier in this section.

Contrary to the framing of smartness as enhancement of a city's technological infrastructure, sometimes the smartest thing to do is not a technical solution. New York City is one of the world's most iconic self-proclaimed smart cities. The city has dedicated tens of millions of dollars to becoming a smart city. At the same time it still suffers from a racially divided school system where Black students, composing 24% of the district, account for 61% of the expulsions and Hispanic students, 41% of the district, account for only 11% of the gifted and talented programs (Groeger, Waldman, and Eads 2018). Meanwhile, prior to the devastation of COVID-19, the city already suffered from a lack of affordable housing and homelessness. Due to COVID-19 it is estimated that homelessness in New York City will dramatically increase (Chadha 2020). Would a smart city not solve these problems first, or at least simultaneously? Might a smart city be one equipped to equitably care for its residents in a pandemic? Increased dependence on the internet during the pandemic has also exposed that NYC still has yet to equitably provide internet access, as promised in 2016 when they were awarded the title of world's best smart city (Media Contact NYC Press Office 2020). This demonstrates that even tech-driven "smart" benefits are often measured only in improvements for the already privileged.

Activists have long shared ideas, informed by the lived experiences and wisdom of citizens, of what would make a city like New York great. Jane Jacobs famously argued as far back as 1961 that roads and skyrises were killing the life of American cities. She saw walkability and diversity of age and purpose of buildings as aspects that kept cities both safe and lively due to eyes on the ground (Jacobs 1961). Although she left out race from her analysis, failing to recognize that apartments might be the only affordable option for displaced groups, she later noted that ghettos were only possible through purposeful government practices such as redlining (Desrochers 2007, 128–129). The most valuable aspects of cities cannot readily be packaged in corporate technical solutions. Clearly, New York City has not tried to be smart for everyone or valued everyone's smartness. Redlining (even if no longer legal) still dictates who is worthy of a good home. And, intentionally or not, gentrification becomes a form of housing eugenics. Some are snipped away from the community, but the value of the neighborhood goes up without them.

Saskia Sassen has a term for what really builds the city, "cityness": the connections and moments of citizens—the art, events, culture—that city residents provide (Sassen 2005). As Sassen explains, cities and their neighborhoods manage to outlive governments and big corporations, even if they become populated with varying people (Raje 2016). The people who make up the city create the city. To ignore the experiential knowledge of residents of a city, or, worse, restrain residents from free engagement with it by imposing rigid technical infrastructures, limits the creativity and advancement of the metropolis. Some ambitions of the smart city may indeed increase the longevity of the city and improve daily life, such as investing in sustainable transportation (Bamwesigye and Hlavackova 2019) and e-government solutions for improved organization and accessibility (Oliveira, Oliver and Ramalhinho 2020). However, adding tech does not guarantee that an idea is smart. Smartness also lies in the experiential and practical knowledge of those deemed unimportant to public space.

5. Unintelligent Decisions in the Name of "Smartness"

Unintelligence has been built into cities in the name of smartness. Redlining, traffic filled city roads, gentrification, and hostile infrastructure are all government implemented policies that were smart only for a privileged subset of the population, thus diminishing the overall quality of city life. Willful ignorance too plays a role in city design. There are set recommendations on how to build disability-compliant cities such as the United Nation's 2016 report, "Good Practices of Accessible Urban Development" (Ito et al. 2016), but investment in social equity does not bring as much enthusiasm as investment in business or providing direct improvement to those the city wishes to support. Making cities accessible is not an act of charity, rather, refusing to do so is unjust exclusion (Mintz 2021).The stupidity in many urban transportation policies has become so apparent that some European cities are taking a step back from prior corporate-driven infrastructure and giving greater control back to pedestrians. For example Utrecht, a city 30 minutes east of Amsterdam in the Netherlands finally in 2020 completed removing the city's main ring roadway and restoring its historic canal (Wagenbuur 2020).

It is not difficult to envision how leaning too hard in the direction of building smart infrastructure, as narrowly defined today, could cause a similar cycle where in a century we are in desperate need to "de-smart-ify" our cities. Replacing live personnel with digital kiosks has already become an issue for blind people who cannot read flat screen buttons with their fingers.[5] While a perfectly efficient smart city might be the dream, the failure of the connected city might become a nightmare. The smart city might become the hackable city or the commercially owned and controlled city. The same network providers quickly working to build 5G are the same that advertise 5G-dependent smart city solutions. What business model could be more profitable than having a whole city dependent on a corporation's network, solution, data management, and upkeep? When one zooms closely into the dreams of the prototypical smart city, the day to day improvements for citizens seem negligible or even disappear.

The typical smart city is not promising to eradicate homelessness, boost childhood education levels, increase democratic involvement, equalize gender rights, eliminate racism, and bring the rest of the world up with it along the way. This is not because these aims are unattainable. Indeed, smart cities promote equally if not more ambitious goals, such as making the city more sustainable and safer, usually by means of increased surveillance and energy dependence. How can a city be made instantly safer while at the same time being more reliant on new, still-developing technology? How can a city become more sustainable while at the same time becoming more energy reliant upon power-hungry kiosks, cameras, lights, and data storage?

Meanwhile, truly smart activities are often repressed by the government, while unintelligent behavior is encouraged under the guise of resilience. Consider intelligence as having a value, knowledge of the relevant domain, and then taking steps to increase the value based on that knowledge. If you are repressed it is therefore smart to protest. The value of freedom, knowledge of being treated incorrectly, and awareness that protest might lead to peaceful change, makes protesting a smart action to take. Sleeping outside or in a tent is smart if you value shelter and safety and have nowhere to go. Peeing on the street makes sense if you need to relieve yourself and do not have a place to do so. Selling drugs is logical if your school system is malfunctioning, you have bills that are impossible to pay, and have no other reliable way to build a livable income.

Conversely, unintelligence is defined as being aware of a problem, having the ability to fix it and not doing anything to exercise that ability. San Francisco and the greater Bay Area are home to some of the world's wealthiest inhabitants and most successful information technology companies, yet struggle to manage their large homeless population, whose desperation and suffering is plainly visible to all. The Bay Area is one of the wealthiest regions in the world where tech companies, dissatisfied with public transportation options for their employees, invest in private charter buses with Wi-Fi and neighborhood pickup and drop off to the office headquarters. Yet when those who struggle due to being neglected by the government do manage to survive, they are lauded as resilient and praised for their self-advocacy, diverting the focus from the failures of government.

The 2014 EU report on smart cities claimed that smart cities are needed in order to have smart citizens; for example, if there is pollution in an area people cannot readily move, however they can use sensors to plan different commuting routes or even plant trees (Manville et al.). Yet it is doubtful that people in such a situation need the smart sensor network to know that their air is unclean. Moreover, failure to bring clean air to citizens is a fundamental failure of a smart city, not the responsibility of smart people to avoid.

In the Netherlands, there is a quite smart development of urinals that pop out of the ground at night around pub areas. While this is a smart solution, it is notable that there are limited options for the 50% of the population unable to use urinals.[6] Perhaps one could argue that women are less likely to pee on the street and therefore less of a problem. But if one thinks of the needs of women as opposed to the needs of the city then it becomes clear that women might need a late-night option even more than men who could somewhat discreetly relieve themselves anywhere. Smart city design choices can easily and comfortably fit this scheme, where everyone benefits from a small subset getting the most improvements. Less pee is on the street, but the main beneficiaries are people who already could have waited for the bathroom like everyone else. The smart city is advertised with a certain set of values, and then focuses on the subset that most closely reflects corporate and governmental interests (or those who happen to be in such demographics). Those consistently denied updates to make harmful infrastructure more livable are gaslit to believe that the smart city goals aligns with their interests because "smart" improvements help "everyone."

6. Technology Is Not Inherently "Smart"

Technology will not make us smarter. At least not technology by itself. The novel coronavirus pandemic that swelled in 2020, followed swiftly by racial injustice protests sweeping from the United States across the globe brought to light ongoing concerns about the unintelligence of smart city technology. They highlighted the misalignment of government, corporate, and community values. Policing tools such as license plate tracking and facial recognition are universally applied, but the technology works by segregating suspicious individuals from those presumed harmless to society. The citizens deemed "normal" or "belonging" do not need to worry about surveillance because the infringement is not hostile to them. Conversely, a tracking application for a highly contagious virus implicates everyone. Privilege does not guarantee immunity, meaning that even with possible health concerns at stake, infection surveillance measures must appeal to the most valued citizens and therefore may be subject to more scrutiny.

The pandemic showed that differing political and cultural regimes can define "smart" technological solutions in their own ways. Tech companies that wanted to build contact tracing apps in the Netherlands had to submit proposals to be reviewed by ethicists, scientists, and government officials (Loohuos 2020). China rolled out individual QR codes to mandate who could be in public space (Mozur, Zhong, and Krolik 2020). Korea used its preexisting smart city infrastructure to track individuals known to have COVID-19 with surveillance camera footage and used phone GPS data to make sure quarantine practices are followed (Wray 2020). Testing and care also differed across countries. Some chose to test all residents in order to trace the spread of the virus. In the

United States tests were expensive and administered disproportionately to those who could afford it, even though the people most likely to die were low-income people of color (Oppel et al. 2020).

Interestingly, it seems to be those most empowered who find it easiest to evade surveillance and subjugation to the corporate and governmental whims of the smart city. San Francisco, home to multitudes of high-paid tech workers, was the first major American city to ban facial recognition technology (Conger, Fausset and Kovaleski 2019). These concerns about facial recognition were amplified in the United States with the Black Lives Matter protests over unjust policing. Whether out of good will or for good publicity, companies such as IBM, Amazon, and Microsoft, decided to stop using and selling facial recognition technology to US police forces, at least temporarily (Hale 2020). In this case, the corporations sided with an increasingly popular movement against an unjust government system, but it is unclear whether this has any lasting significance for undervalued citizens. Will the restraint on sale of facial recognition tools to police forces persist if and when active support for the BLM movement ebbs among the dominant white majority, as it frequently has?

Smart government is potentially dangerous when a privileged elite uneducated on the needs of the many (such as software engineers) build supposedly apolitical decision-making machines. Fairness and neutrality cannot simply self-manifest in a subjective and biased system. Supervised machine learning still needs to be fed examples to learn from, or as often is the case, mistakes to repeat. Groups can be subjected to algorithmic violence through tools such as Market Value Analysis (MVA), a system that predicts what neighborhoods will be valuable and therefore worth investment. MVA has guided city development across the United States since 2001, potentially creating a new form of digitally determined redlining (Safransky 2019, 201). The smartness of crime algorithms is also called into question when they are designed with fixed biases (such as flagging licenses as more likely to be connected to crime if they are from certain countries, or the model of the car is from a certain year) or trained upon datasets rife with unjust historical bias. For example, historical over-policing of minoritized or impoverished areas means algorithms trained on that data will predict more crime in those areas, making it more likely that police will be sent to that area and find cases of crime. Sociologist Ruha Benjamin has noted that such tools become self-fulfilling prophecies leading crime prediction algorithms into runaway feedback loops of crime *production* (Benjamin 2019, 83). Technology built upon broken systems without the goal of fixing the underlying conditions cannot improve a city. It takes smart social action and not just data to solve these pre-existing conditions.

Furthermore, technology breaks. Resilience has been another popular catchphrase for cities. A smart city is able to bounce back. Yet often the pressure of being able to bounce back is put on the individual. The smart city gets to try and fail, while the citizens, especially the most at risk, must be resilient and endure the harms caused by these failures. IBM released a white paper outlining 17 vulnerabilities that they found testing smart city security. Some of the most basic hacking techniques that occurred in multiple cases were due to public default passwords, easy authentication bypass, and

SQL injection (Crowley 2018). The hacked smart city evokes hijacked public vehicles such as busses, trains, trams, and police cars, long term grid malfunction, false emergency alerts, and killed agriculture. Theoretically whatever is built and connected can be hacked. A city hack already occurred in Ukraine in 2017 when the power grid was shut down by hackers running a test on the system (Greenberg 2017).

Ironically, it is the corporations that are selling the solutions to poor technology. Subscribing to a smart city plan backed by a large tech company grants peace of mind because they ensure a centralized solution with the industry's leading security. However, sacrificing control over the city to a corporation might not be worth it. What happens if the government grows unhappy with a service that has been integrated into all aspects of government? People also forget the impact of those who build and maintain the technology. Rarely do engineers of the smart city reflect the diversity of the citizens. While they may have technical expertise it is unlikely for them to have the intelligence needed (the experiential knowledge, skill in navigating, or even awareness of regional and minority issues.) Corporations and governments tend to direct questions in this realm toward the fact that they encourage public-private partnerships. This returns us to the original question: is the technology developed in such partnerships likely to be the most appropriate and effective solution to the challenges that citizens know? Will it be installed regardless of citizen need? Will its governance be guided by the knowledge that citizens possess?

There is a growing call for contestable infrastructure: ICT that allows the dissatisfied to talk back, and the pleased to share what features they enjoy. If the goal is really to build a smart city that is dynamic and citizen-driven, it should also be flexible and made so that it fits the needs of the full community. Technology alone is not a solution. Moreover, an ICT tool is not inherently better than any other social initiative, and both are only as smart as their guidance and regulation by the irreplaceable knowledge of citizens.

7. Conclusion

The smartness of the smart city has been interrogated. It has been found that there is a high risk that the smart city will not be smart for everyone, nor reflective of the knowledge held by smart citizens. Given the ICT-driven interests of smart cities and the historical tendency of city and corporate leaders to ignore the humanity of subsets of the population, it is possible for cities to become "smart" while still failing a substantial subset of their constituents. In order to make smart cities smart for all, it first must be clear that is the intention. Next, it is essential that the experiential knowledge of citizens not just be seen as important, but valued and necessary. Behavior that goes against the ICT-based definitions of the smart city, may add intelligence to the city if the goal is understanding and supporting the lives of citizens and their opportunities for growth, rather than repressing their agency to maintain the status quo.

In this age of smart cities, cities are recognizing their own deficiencies—that they may lack the newest technology or the most sustainable infrastructure. In order to compete, they are investing money in upgrading by bringing in new sources of knowledge in the form of sensors. Yet this smart city phase is also an opportunity to recognize other types of failures and build institutional solutions that will improve education, healthcare, and sustainability, without becoming reliant on the next new shiny tool a company wishes the city to buy. This is not to say that there are not good and needed ICT solutions. Improving electrical grids, sewer systems, and connectivity across socioeconomic groups are investments likely to pay substantial dividends in the future. Housing unnecessary data in city servers that drain electronic and energy resources does not seem as useful. Collecting data on in-need communities, instead of financing solutions that the community likely could identify and communicate themselves, seems like a waste of money.

A true smart city uses technology flexibly and cannot be crashed by a failure in one central system. A true smart city might first make all the public resources disability friendly. It might provide the resources needed to make those likely to be attacked for their identity feel safe. A true smart city might permit some discomfort and disruption to be experienced by those who the city *currently* works for, because it will take the needs of the invisible seriously and bring the necessary supportive infrastructure to life.

There is often a feeling that the norm (whether it is wrong or right) is the baseline, making other possible futures strange, uncomfortable or unlikely. For example, white men are unlikely to strongly identify with being white or male even though that identity strongly shapes how they move, perceive, and interact with the world. Able-bodied people are unlikely to identify as able-bodied even though being able to see, walk, and hear defines their existence. In order to critically shape future cities, the "norm" must be re-understood, not as the logical status quo, but instead as a prejudiced infrastructure built upon the erasure of groups stripped from positions of power. One cannot help others if they are so blinded by their normalized privilege that they cannot see how others could struggle to cope with the same infrastructures that support them. The smart city needs to be reframed. A city, with or without technology, that ignores the demands of those most in need is an unintelligent city at best, and a hostile city at worst.

Acknowledgements

The research presented in this paper has been funded by the Netherlands Organisation for Scientific Research (project: Bridging Data in the built environment (BRIDE); project number: CISC.CC.018).

Notes

1. Throughout this chapter, "citizen" refers to city residents broadly and does not imply national citizenship.
2. County technology adoption rates were used for this deduction. Many successful smart cities are located in nations that have been early tech adopters. https://ourworldindata.org/technology-adoption

3. A study conducted by World Enabled found that “only 4% [of 1200 digital city projects from six separate agencies] specifically referred to people with disabilities and older age groups.” https://news.trust.org/item/20191121165730-w14i4/
4. Until 2013 Sweden also required individuals wishing to have their sex legally reassigned to undergo sterilization. These individuals, along with those who underwent coercive sterilization under the 1934 and 1942 sterilization acts, are now eligible for government compensation. Sweden is not alone. More than a dozen European countries had similar requirements in 2017 when the European Human Rights Court deemed such laws unethical. Such laws span beyond Europe, in 2019 a transman lost his case to Japanese parliament to be recognized as male without sterilization. Unsurprisingly, Japan too has a history of forced sterilization. See: https://www.thelocal.se/20130111/45550, https://www.nytimes.com/2017/04/12/world/europe/european-court-strikes-down-required-sterilization-for-transgender-people.html, https://thediplomat.com/2019/02/japans-supreme-court-upholds-surgery-as-necessary-step-for-official-gender-change/
5. There are kiosk options for people with vision impairments, but they are not yet universally implemented.
6. Amsterdam unveiled the first pop-up toilet for those unable to use a urinal in 2016. They are uncommon possibly due to the design. See: https://www.dutchnews.nl/news/2016/03/87534-2/

   It has also been noted that those without penises may also use the retractable public outdoor urinals if they carry around a device created by Dutch inventor Moon Zijp called a “plastuit” in Dutch (imagine a funnel placed between the legs).

References

Adam, Alison. 1998. *Artificial Knowing: Gender and the Thinking Machine*. Florence, KY: Routledge.

Adams, David Wallace. 1997. *Education for Extinction: American Indians and the Boarding School Experience, 1875–1928*. Lawrence, Kansas: University of Kansas Press.

Bamwesigye, Dastan, and Petra Hlavackova. 2019. “Analysis of Sustainable Transport for Smart Cities.” *Sustainability* 11 (7): 2140.

Benjamin, Ruha. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons.

Brown v. Board of Education. 1953. 347 U.S. 483 (Supreme Court of the United States).

Cascone, Sarah. 2020. *The Organization in Charge of Rebuilding Notre Dame Must Be More Transparent About Its Use of Donations, a French Court Says*. October 2. Accessed December 22, 2020. https://news.artnet.com/art-world/french-court-auditors-rules-notre-dame-donations-1912604.

Cave, Stephen. 2020. “The Problem with Intelligence: Its Value-Laden History and the Future of AI.” *AIES ‘20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 29–35.

Chadha, Janaki. 2020. “New York Is Facing a Potential Explosion in Homelessness.” *Politico*. June 30. Accessed July 30, 2020. https://www.politico.com/states/new-york/city-hall/story/2020/06/29/new-york-is-facing-a-potential-explosion-in-homelessness-1296100.

Conger, Kate, Richard Fausset, and Serge F. Kovaleski. 2019. “San Francisco Bans Facial Recognition Technology.” *The New York Times*. May 14. Accessed July 31, 2020. https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.

Crowley, Daniel. 2018. "How to Outsmart the Smart City." *Security Intelligence*. August 9. Accessed July 30, 2020. https://securityintelligence.com/outsmarting-the-smart-city.

Cumming, Daniel G. 2018. "Black Gold, White Power: Mapping Oil, Real Estate, and Racial Segregation in the Los Angeles Basin, 1900–1939." *Engaging Science, Technology, and Society* 4: 85–110.

de Fine Licht, Karl. 2017. "Hostile Urban Architecture: A Critical Discussion of the Seemingly Offensive Art of Keeping People Away." *Etikk I Praksis:Nordic Journal of Applied Ethics* 11 (2): 27–44.

Desrochers, Pierre. 2007. "The Death and Life of a Reluctant Urban Icon." *Journal of Libertarian Studies* 21 (3): 115–136.

Doctoroff, Daniel L. 2020. *Why We're No Longer Pursuing the Quayside Project—and What's Next for Sidewalk Labs*. Side Walk Talk. May 7. Accessed July 27, 2020. https://medium.com/sidewalk-talk/why-were-no-longer-pursuing-the-quayside-project-and-what-s-next-for-sidewalk-labs-9a61de3fee3a.

European Commission. n.d. *Smart Cities*. Accessed July 23, 2019. https://ec.europa.eu/info/eu-regional-and-urban-development/topics/cities-and-urban-development/city-initiatives/smart-cities_en.

Galdon-Clavell, Gemma. 2013. "(Not So) Smart Cities? The Drivers, Impact and Risks of Surveillance-Enabled Smart Environments." *Science and Public Policy* 40: 717–723.

Giles, Chris. 2018. "African 'Smart Cities:' A High-Tech Solution to Overpopulated Megacities?" *CNN*. April 9. Accessed July 31, 2020. https://edition.cnn.com/2017/12/12/africa/africa-new-smart-cities/index.html.

Government of Sweden. 1992. *Sterilization Issues in Sweden 1935–1975: Financial Compensation*. Government Report, SOU.

Greenberg, Andy. 2017. " 'Crash Override': The Malware That Took Down a Power Grid." *Wired*. June 12. Accessed July 31, 2020. https://www.wired.com/story/crash-override-malware

Grodin, Michael A., Erin L. Miller, and Johnathan I. Kelly. 2018. "The Nazi Physicians as Leaders in Eugenics and "Euthanasia": Lessons for Today." *American Journal of Public Health* 108 (1): 53–57.

Groeger, Lena V., Annie Waldman, and David Eads. 2018. "Miseducation: Is There Racial Inequality at Your School." *ProPublica*. October 16. Accessed July 28, 2020. https://projects.propublica.org/miseducation.

Hale, Kori. 2020. "Amazon, Microsoft & IBM Slightly Social Distancing from the $8 Billion Facial Recognition Market." *Forbes*. June 15. Accessed July 31, 2020. https://www.forbes.com/sites/korihale/2020/06/15/amazon-microsoft--ibm-slightly-social-distancing-from-the-8-billion-facial-recognition-market/#2f20c8164a9a.

Ito, Akiko, Guozhong Zhang, Maria Martinho, Robert Venne, Miranda Fajerman, Julie Pewitt, Talin Avades, and Claire Odom. 2016. *Good Practices of Accessible Urban Development*. United Nations Department of Economic and Social Affairs.

Jacobs, Jane. 1961. *The Death and Life of Great American Cities*. New York: Random House.

Joerges, Bernward. 1999. "Do Politics Have Artefacts?" *Social Studies of Science* 29 (3): 411–431.

Kestling, Robert W. 1998. "Blacks Under the Swastika: A Research Note." *The Journal of Negro History* 83 (1): 84–99.

Khare, Vineet. 2019. "India Election 2019: Have 100 'smart cities' been built?" *BBC News*. March 21. Accessed July 31, 2020. https://www.bbc.com/news/world-asia-india-47025472.

Kochan, Thomas, Katerina Bezrukova, Robin Ely, Susan Jackson, Aparna Joshi, Karen Jehn, Jonathan Leonard, David Levine, and David Thomas. 2003. "The Effects of Diversity on

Business Performance: Report of the Diversity Research Network." *Human Resource Management* 42: 3–21.

Kukla, Rebecca. 2006. "Objectivity and Perspective in Empirical Knowledge." *Episteme* 3 (1-2): 80–95.

Lombardo, Paul A. 2010. *Three Generations, No Imbeciles Eugenics, the Supreme Court, and Buck v. Bell*. John Hopkins University Press.

Loohuos, Kim. 2020. "Coronavirus: Dutch Covid-19 Tracking App Stirs National Debate." *Computer Weekly*. April 24. Accessed July 30, 2020. https://www.computerweekly.com/news/252482131/Coronavirus-Dutch-Covid-19-tracking-app-stirs-national-debate.

Manville, Catriona, Gavin Cochrane, Jonathan Cave, Jeremy Millard, Jimmy Kevin Penderson, Andrea Liebe, Matthias Wissner, Roel Massink, Bas Kotterink, and Rasmus Kåre Thaarup. 2014. *Mapping Smart Cities in the EU*. European Union.

McNeill, William. 1999. "How the Potato Changed the World's History." *Social Research* 66 (1): 67–83.

Media Contact NYC Press Office. 2020. "Mayor de Blasio and Taskforce on Racial Inclusion and Equity Announce Accelerated Internet Master Plan to Support Communities Hardest-Hit by COVID-19." *NYC.gov*. July 7. Accessed July 29, 2020. https://www1.nyc.gov/office-of-the-mayor/news/499-20/mayor-de-blasio-taskforce-racial-inclusion- equity-accelerated-internet-master.

Mintz, Kevin Todd. 2021. "Universally Designed Urban Environments: 'A Mindless Abuse of the Ideal of Equality' or a Matter of Social Justice?" In *Technology and the City: Towards a Philosophy of Urban Technologies*, edited by Michael Nagenborg, Taylor Stone, Margoth González Woge and Pieter E. Vermaas. Springer International Publishing.

Mozur, Paul, Raymond Zhong, and Aaron Krolik. 2020. "In Coronavirus Fight, China Gives Citizens a Color Code, with Red Flags." *The New York Times*. March 1. Accessed July 31, 2020. https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html.

New York City Hall Press Office. 2016. *The Official Website of the City of New York*. November 28. Accessed July 23, 2020. https://www1.nyc.gov/office-of-the-mayor/news/909-16/new-york-named-2016-best-smart-city-nyc-host-2017-international-conference-urban.

NYC Mayor's Office of the Chief Technology Officer. n.d. Accessed July 23, 2020. https://iot.cityofnewyork.us/data-management/.

NYC Open Data. 2019. "The Next Decade of Open Data 2019: Open Data for All Report." New York.

Oliveira, Thays A., Miquel Oliver, and Helena Ramalhinho. 2020. "Challenges for Connecting Citizens and Smart Cities: ICT, E-Governance, and Blockchain." *Sustainability*.12 (7): 2926.

Oppel, Richard, Robert Gebeloff, K.K. Rebecca Lai, Will Wright, and Mitch Smith. 2020. "The Fullest Look Yet at the Racial Inequity of Coronavirus." *The New York Times*. July 5. Accessed July 31, 2020. https://www.nytimes.com/interactive/2020/07/05/us/coronavirus-latinos-african-americans-cdc-data.html.

Oved, Marco Chown. 2019. "Google's Sidewalk Labs Plans Massive Expansion to Waterfront Vision." *Toronto Star*, February 14.

Paulas, Rick. 2019. "Photos of the Most Egregious 'Anti-Homeless' Architecture." *Vice Media Group*. January 25. Accessed July 31, 2020. https://www.vice.com/en_us/article/kzm53n/photos-of-the-most-egregious-anti-homeless-architecture.

Purbrick, Martin. 2019. "A Report of the 2019 Hong Kong Protests." *Asian Affairs* 50 (4): 465–487.

Raje, Aparna Piramal. 2016. "Redefining Notions of Urban Intelligence." *Live Mint*. June 29. Accessed July 30, 2020. https://www.livemint.com/Specials/m21w1rzMM8KpbE9KO1iFVK/Redefining-notions-of-urban-intelligence.html.

Rasmussen, Wayne D., George Edwin Fussell, Kenneth Mellanby, Kusum Nair, George Ordish, Gary W. Crawford, and Alic William Gray. 2020. "Origins of Agriculture." *Encyclopedia Britannica*. February 4. Accessed September 2020. https://www.britannica.com/topic/agriculture/The-Americas.

Rosatia, Umberto, and Sergio Contia. 2016. "What Is a Smart City Project? An Urban Model or a Corporate Business Plan?" *Procedia Social and Behavioral Sciences* 233: 968–973.

Rosenberger, Robert. 2020. "On Hostile Design: Theoretical and Empirical Prospects." *Urban Studies* 57 (4): 883–893.

Safransky, Sara. 2019. "Geographies of Algorithmic Violence Redlining the Smart City." *International Journal of Urban and Regional Research* (44)2: 200–218

Sassen, Saskia. 2005. "Cityness in the Urban Age." *Urban Age*.

Schreiner, Clara. 2016. *International Case Studies of Smart Cities Rio de Janeiro*, *Brazil*. Inter-American Development Bank.

Steup, Matthias, and Ram Neta. 2020. "Epistemology." *The Stanford Encyclopedia of Philosophy*. June 21.https://plato.stanford.edu/archives/fall2020/entries/epistemology/

Ting, Tin-yuet. 2020. "From 'be water' to 'be fire': nascent smart mob and networked protests in Hong Kong." *Social Movement Studies* 19 (3): 362–368.

Wagenbuur, Mark. 2020. "Utrecht Corrects a Historic Urban Design Mistake." *Bicycle Dutch*. Accessed December 23, 2020. https://bicycledutch.wordpress.com/2020/09/16/utrecht-corrects-a-historic-urban-design-mistake/.

Weindling, Paul. 1989. *Health, Race, of German Politics Between National Unification and Nazism, 1870–1945*. Cambridge, UK: Cambridge University Press.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109, no. 1: 121–136.

Wray, Sarah. 2020. "South Korea to Step-up Online Coronavirus Tracking." *Smart Cities World*. March 12. Accessed July 31, 2020. https://www.smartcitiesworld.net/news/news/south-korea-to-step-up-online-coronavirus-tracking-5109.

Yang, Yuan. 2020. "Why Hong Kong Protesters Fear the City's 'Smart Lamp Posts.'" *Financial Times*. January 8. Accessed July 27, 2020. https://www.ft.com/content/f0300b66-30dd-11ea-9703-eea0cae3f0de.

PART III

TECHNOLOGY, POWER, AND POLITICS

CHAPTER 10

PHILOSOPHY OF TECHNOLOGY AS POLITICS

ADAM BRIGGLE

1. Philosophy as Technology

In his classic essay, Langdon Winner asks "Do Artifacts Have Politics?" (Winner 1980). His answer: yes. He acknowledges that technologies are shaped by social and economic forces, but he argues that artifacts are politically significant in their own right. They are not neutral, pliable things wholly determined by interest groups, class struggles, or elite power. No, they also transform people and give rise to new social dynamics and "forms of life" (Winner 2014). Later theorists would similarly describe technologies as "actants" or "mediators" that play active roles in shaping society (e.g., Latour 2007; Verbeek 2005).

Here, I ask the meta-version of that question: Does *the philosophy of* artifacts have politics? My answer: yes. Indeed, the philosophy of technology has politics precisely because it, too, is an artifact. It is a technique that embodies forms of power and authority, operates by its own imperatives, and imposes its own ideology on its practitioners who become its spokespeople. I am not making the trivial claim that philosophers of technology have their own political opinions or even that some of them philosophize *about* politics. Rather, I am calling attention to the techniques of philosophizing and their unavoidable political dimensions. My goal is to get those who think about the non-neutrality of artifacts to consider the non-neutrality of their thinking.

The technique that I have in mind is the academic discipline. Now, philosophy of technology was not originally its own sub-discipline. As Carl Mitcham (1994a) points out, several early contributors were engineers and entrepreneurs, sometimes not academics at all. But across the twentieth century, philosophy became an academic discipline (see Bordogna 2008)—carrying along with it the philosophy of technology and other sub-disciplines. Standard histories of philosophy focus only on the conversations around various ideas about, say, truth, justice, beauty and the good life. Yet these conversations have not happened in vacuums—they have occurred in shifting social and institutional contexts. One of the most important organizational developments in the history of philosophy was the academic discipline with its politics of peer-review, specialization, professionalization, and expertise. This development occurred roughly between 1865 and 1920 and was part and parcel of the advent of the modern research university (see Veysey 1979).

Because this institutional history of philosophy has mostly been overlooked, the discipline forms the dominant and largely unexamined technology of philosophy and, thus, its politics. Winner talks about the imperative of the atom bomb—how it demands a centralized, hierarchical, secretive politics. I want to consider the institutional imperatives of the academic discipline and how they also demand certain arrangements of power and authority.

Most important, they compel the philosopher to become an expert who masters a discrete domain with its own language, standards of success, and specialized concerns. This narrows the scope of their audience and their accountability—they are turned inward to talk only to other philosophers. According to disciplinary politics, if philosophers win the approval of other philosophers, then they are successful. Yet philosophers of technology in particular could help engineers, policymakers, consumers, and citizens to make more reflective and critical choices about technologies. There is a tremendous need for this help. Fulfilling that need should be the prime directive for philosophers of technology. However, under the politics of the discipline this is, at best, something tacked onto the side of their core responsibilities.

2. Philosophy of the Discipline

Across the twentieth century, philosophers paid increasingly careful attention to technology. Yet the type of technology (understood as knowledge or technique) that most profoundly impacted their own work largely went un-theorized. By the end of that century, philosophy had become a discipline with undergraduate majors, graduate programs, and dozens if not hundreds of specialized journals. Knowledge production ramped up impressively, but precious little of that activity turned inward to ask how the disciplining of philosophy changed philosophy itself (see Frodeman and Briggle 2016b). This is true across the West and indeed increasingly also in China and other places adopting modern modes of knowledge production.

In their 2000 compendium on the philosophy of technology, Peter Kroes and Anthonie Meijers describe the field as “A Discipline in Search of Its Identity.” But a discipline is already an identity. That’s another way to say that it has politics. It is not a blank canvas on which one paints or an empty box into which one places an identity. It is common for philosophers of technology to assume that the discipline is simply the condition for the possibility of thinking and not itself a mediator of their thinking. As I argue elsewhere (Briggle 2016), this is a curious blind spot for a community so attuned to the ways techniques mediate cognition and social activities.

Let us, then, philosophize about the artifact of the discipline. I will do so briefly in three ways: first with the assistance of Martin Heidegger's essay "The Question Concerning Technology" (1977), then with some help from interdisciplinarity studies, and then with the aid of Ivan Illich.

Heidegger argues that the essence of technology doesn't have anything to do with its material. Rather, technology is a way of revealing or bringing forth the world. Modern technology is what he calls a "challenging revealing" where everything is ordered to stand by and be on reserve for the sake of ever more efficient productivity. He calls this kind of revealing or order *Gestell*. The tricky thing about *Gestell* is that it drives out every other possible way of revealing, and it does so by concealing revealing itself. That is, it is taken to be simply reality—not one way of ordering things but simply *the* order of things.

This is also the case with the discipline and the way in which it tends to disappear from view to become simply *the* view, that is, the way through which the world is viewed. The discipline, too, is a kind of "challenging revealing" that demands every topic to show itself according to the discipline's terms. It is the hammer that makes nails of all the world. Daniel Callahan (1973) calls this "disciplinary reductionism," the penchant for distilling out of a complex problem one issue, which is then labeled *the* issue. Naturally, the issue fits comfortably within the usual way the discipline talks and so it is treated through its normal mechanics. It becomes fodder for the productive churn, and another publication results. But, as Callahan notes, not a publication of any use to the people actually struggling with the original problem in its "unreduced" form.

I have been guilty of this kind of behavior. I once published an academic article about a controversial windfarm. My argument was that the debate hinged on aesthetics. However, in crafting a suitably clean or manageable philosophical argument, I scrubbed out the politics and economics of the controversy. The end result was a rather abstract treatise about the aesthetics of the windfarm that didn't relate to any of the stakeholders' actual concerns as they existed within the complex issue itself.

Now to turn to some of the literature from interdisciplinary studies. As others have noted (see Frodeman, Klein, and Pacheco 2015), academic disciplines are not (merely) epistemic categories. In other words, they do not carve nature at the joints as if there was in reality distinct domains corresponding to biology, political science, chemistry, geology, economics, physics, history, etc. Rather, the disciplines are political units in several senses of that term. Perhaps most obviously, they are political-economic units that shape funding streams via student credit hours, tuition dollars, and administrative and state distributions. I will briefly consider three other important political dimensions of the discipline.

First, disciplines establish lines of power and authority both as a matter of internal institutional dynamics and outward-facing social roles. Internally, disciplinary peers serve as gatekeepers to accredit the next generation and to police acceptable speech, standards of rigor, and indicators of success. They define, in other words, what shall count as "real" research or scholarship. Even as more and more knowledge is *produced* outside of the university, the discipline remains the primary unit for *legitimizing or accrediting* new claims to knowledge. Externally, disciplines set up practitioners as the experts with specialized knowledge in their domains. The politics of the philosophical discipline, then, is philosopher-as-expert, where the expert masters one region of knowledge supposedly distinct from other regions of knowledge. So one way to see philosophy *as* technology is to note how it aped the turn toward specialization and expertise in the natural sciences and engineering. Soon after the technical and natural sciences developed into disciplines, philosophy and the other humanities followed suit. They adopted the same identity (experts) and mode of specialized knowledge production.

Second, disciplines compel academics to adapt to their imperative to "publish or perish." In other words, the discipline is a productivist technique premised on the endless churn of peer-sanctioned knowledge. Disciplines are often described as the units of intellectual autonomy or self-governance, which is true in some ways. But as Steven Turner (2000) notes, they are better described as *autotelic* or self-justifying: they are the consumers of their own productions and there is no governor on that process. In philosophy, this leads to insular discourse of philosophers talking only to other philosophers. Escaping this imperative is difficult, as Wolfgang Krohn argues: "Whatever drives people into highly complex interdisciplinary projects—curiosity, social responsibility, or money—the need of manageable objects and presentable results in their reference community drives them out again" (Krohn 2010, 32).

Winner talks about the imperatives of certain technologies as "the moral claims of practical necessity" (1980, 132). Maybe liberty, equality, and participation are good things for the railyard worker, but that's no way to run a railroad. The trains must arrive on time. Values not reflected in the imperative appear obsolete, irrelevant, or foolishly idealistic. Analogously, as Krohn argues, engaging directly with real-world issues is a good thing for philosophers ... but that's no way to run a discipline! One must produce peer-reviewed publications, those objects for the reference community of disciplinary peers. To help, say, a group of prisoners understand and protest their unfair working conditions is foolish and "not real philosophy." Yet there is distinctively philosophical help to be offered in such conditions—for example, one can analyze arguments, clarify values, supply concepts for characterizing power dynamics, and much more.

Third, the discipline carries its own political ideology, which is basically libertarian. Curiosity and freedom of inquiry are the reigning ideals. This is especially true in the American context (the Code of Conduct for the American Philosophical Association prioritizes academic freedom of speech above all else). This constitutes the academic as an atomistic thinker with no prior ties or commitments to any larger community. True libertarians, disciplinary philosophers do as they please. They are not obligated to consider the needs or interests of anyone else, and they certainly do not take orders or even suggestions about what they should think about next. Again, this libertarian streak seems to be strongest in the United States. By contrast, European philosophers and philosophers of technology more often break with disciplinary politics to work in collaboration with engineers and other stakeholders.

A third philosophy of the discipline can draw inspiration from Ivan Illich's concept of "radical monopoly" in his book *Tools for Conviviality* (1973). A radical monopoly is dominance by one type of product or production process rather than dominance by a single brand. For example, a road without bike lanes, crosswalks, or sidewalks cedes a radical monopoly to the automobile even if there are many different kinds of automobiles on the road. Pedestrians and bicyclists are excluded. Radical monopolies constitute "a special kind of control" because they impose consumption of a standardized product that only large institutions can provide. In this way, they tend to rule out more informal activities and natural competencies.

The discipline (whether in philosophy or any other area) acts as a radical monopoly in similar ways. Non-disciplinary forms of philosophy exist as marginal or deviant activities in the same way a bicyclist exists on a road without bike lanes in a culture unaccustomed to bicyclists. Things can get risky. Most important, philosophers who have spent significant time working with policymakers or engineers on real-world technology problems often face institutional biases against such work. Whatever "products" they produced (say an advisory report or a set of stakeholder meetings) may not be counted in their evaluation criteria. By establishing the criteria for what shall count as philosophy (usually, publications in peer-reviewed journals) and what shall count as excellent philosophy (usually, a certain collection of top-rated journals), the discipline exerts enormous power. Those who seek to work outside of its established evaluation criteria do indeed take risks similar to those trying to bicycle through a car-centric city.

Too much of the philosophy of technology is obsessed with what are essentially debates about brands: should we be empirically attuned to the wondrous variety of the internet or should we dourly castigate *La Technique*? This is like: should we drive a Ford or a Chevy? Because whether one celebrates technical diversity or lambasts its homogeneity is not nearly as important as *how* one does so and *with whom*. The default answer is that one does so in the lingo of disciplinary jargon with fellow experts. That is to say, the politics of libertarian productivism in philosophy of technology is as dominant as it is invisible.

In other words, the hot debates in the philosophy of technology are about brand names (e.g., post-phenomenological or analytic) and not about product types or production processes. No matter which brand of philosophy one chooses, the politics or the "form of life" is the same. It consists of philosophers producing articles and book chapters for fellow philosophers to read. The material reality is computers in offices, reams of specialized journals, occasional trips to the library, and conference travel. It is a professional life with service obligations to sit on committees, review book manuscripts, etc. It is a career, with a steady pay-check for following one's curiosity as publications become lines on a lengthening CV. In sum, the political force of the discipline dictates who can speak authoritatively, to whom they must speak, and how their speech must sound. The discipline certainly serves its own imperative for knowledge production, but does it offer much by way of social benefit?

3. In Defense of the Discipline

Before critiquing the discipline, I want to come to its defense. Like most technologies, it is a mixed-bag of good and bad properties and impacts. Historically, we could even argue that the discipline was a necessity. In the early twentieth century, with the sciences carving intellectual territory into specialized domains, philosophy faced an existential quandary. What role could it play? In theory, it could have been a synthesizer, a translator, a gadfly, a court jester, etc. But in reality, there was only one choice. Philosophy, too, had to claim some territory as its own. Philosophers had to adapt to the imperative of specialization if they were to survive and reproduce in the evolving knowledge ecosystems of the university. I will argue that in that adaptation, they became deformed, but at least they survived in some shape, thus creating the possibility for further adaptation.

The discipline should not exercise a radical monopoly on the philosophy of technology, but that does not mean there is no place for the discipline. I am a pluralist on this matter. If someone anointed me king of the philosophers, I would spend most of my energy legitimizing and institutionalizing alternative philosophical techniques with their alternative politics. I would, for example, house philosophers in engineering departments to help account for a wider set of values in technical design. And I'd work on developing internships for philosophy students with local governments to help policymakers and staff think through moral dilemmas around, say, land use planning. I'd restructure curricula to help train skills for analyzing real-world problems in real-time in collaboration with various stakeholders. But I would happily allow a sizeable portion of philosophers (maybe a third) to continue doing specialized work.

That's because there are virtues to disciplinarity. I think the main virtue stems from its relationship with that other sense of "discipline" as in "punishment for the sake of correction." The term is related to "disciple" (one who follows another for the purpose of learning), which is rooted in *discipere*, meaning to take apart or to grasp intellectually. Academic disciplines provide the spaces for disciplining the next generation. Even if you favor, as I do, more direct political interventions by philosophers of technology, you must acknowledge the necessity of structured time for young scholars to immerse themselves in the history of philosophy and to practice thinking, which can be a punishing experience. It is hard, as in rigid, as in rigorous. I like the rigor of disciplinary philosophy. I just don't think it is the only, or the most important, kind of rigor.

An old term for disciplinarity is "pure research." Its native habitat is the laboratory, where everything except for the variable of interest can be bracketed and ignored. This is related to that painful meaning of discipline as you must focus or train yourself to follow a path and not wander. As Latour (1993) notes, this is also a defining move of modernity: to purify entirely distinct ontological zones. Yet "pure" sounds also like "irrelevant" or unconnected to practical life. And that cannot be right, because we know that the disciplined thinking (in the non-academic sense of that term) of Isaac Newton led *somehow* to computers and so much more, even if he wasn't tinkering around with anything practical, even if he didn't have any "application" in mind, and indeed even if he couldn't possibly imagine computers.

In his 1945 report, *Science—the Endless Frontier*, Vannevar Bush popularized "basic" as a replacement for "pure" (Bush 1945). "Basic research," he writes, "is performed without thought of practical ends." Yet without basic research, the reservoir of knowledge necessary for improvements in health, security, education, and communication will dry up. "Statistically it is certain that important and highly useful discoveries will result from some fraction of the undertakings in basic science; but the results of any one particular investigation cannot be predicted with accuracy." After all, if they could be predicted, then there wouldn't be the need to do research in the first place.

This argument became the default political contract between science and society. It created a system of accountability premised on disciplinary standards (in the academic sense of that term). Accountability was defined by peer review: scientific research was judged by one's disciplinary peers. If those peers found this research to be good by their standards, then the research was *ipso facto* good for society. Peer review defined the extent of the obligations of the scientist: Just do good research—there is no further need to think about its broader impacts. Society is free to draw from the results as it sees fit, turning research outputs into impacts. How that unfolds is not the business of the researcher. Here, too, philosophy mimicked the sciences by adopting the same legitimizing story about their research—though this story is largely unconscious or at least unspoken.

This is the politics of serendipity (see also Polanyi 1962). In Bush's terms, it is "the free play of free intellects, working on subjects of their own choice, in the manner dictated by their curiosity for exploration of the unknown," that in a happy coincidence yields a healthier, wealthier society. In terms of accounting we only needed to tally peer-reviewed articles. High-quality (i.e., disciplinary, peer-reviewed) research was both the necessary and sufficient condition for highly impactful or socially beneficial research.

There is much to like about this political contract: it shows respect for the wellsprings of creativity, and it wards off the often obtuse nature of political interference. In a defense of disciplinary philosophy of science, Rudolph Carnap argued that "philosophy leads to an improvement in scientific ways of thinking and thereby to a better understanding of all that is going on in the world, both in nature and society; this understanding in turn serves to improve human life" (Carnap 1963, 23–24). So where Bush started with science, Carnap starts with philosophy:

philosophy → physics → chemistry → biology → technology → happiness

In this framing, philosophy would not just model itself on basic science, but also put itself at the very base of basic science. The philosopher's armchair is logically prior to the scientist's lab. Before the scientist can make progress in conquering nature for the relief of man's estate, the philosopher must clear the conceptual ground. This requires retiring from the fray of action into the realm of thought (see Borgmann 1995). In other words, the value or impact of philosophy, like basic science, is indirect and mediated through long chains. It is hard to overestimate, for example, the impact of Francis Bacon, John Locke, or Rene Descartes on the modern world, even though it would be nearly impossible to tease out any direct or simple chains of causality from their philosophical writings to any particular outcome.

In his defense of disciplinary philosophy, Baird Callicott (1999) argues that actions are determined by an "ambient intellectual ether" through which we make sense of our experience. Worldviews, in turn, come to life through the aid of philosophy as it births new "cultural notions and associated norms" (43). Callicott claims that the modern Western worldview is rooted in philosophy from the pre-Socratics onward. And he thinks that disciplinary philosophy is vital today for birthing a post-modern worldview given the ecological destructiveness of the modern technological way of being. Philosophers, he argues, "should not feel compelled to stop thinking, talking, and writing" and "go *do* something" instead (43). Because, their thinking is the most important kind of doing: it is configuring a new worldview.

Chen Changshu (2016) gives a similar defense for the philosophy of technology. He believes that philosophy of technology "has its own independent subject matter, which results in it being a relatively distinct discipline" (11–12). Chen, one of the founders of philosophy of technology in China, argues that "the relationship between technology and philosophy should be considered neither direct nor exceptionally close . . . philosophy and technology act on each other through numerous intermediate links" (2). Philosophy, he claims, does not guide technology. Indeed, it is counterproductive to think that the usefulness of philosophy can be found in any direct social involvement with technology. Rather, the usefulness of philosophy refers to its influence on our "way of thinking," our attitudes, general approach to problems, and our basic conceptions and categories.

4. Why the Discipline Is Bad for Philosophy

Go back to Latour's point about the act of purification being the essence of modernity. That was actually only half of his thesis. The other half is the increasing proliferation of "hybrids" or "monsters" (Latour 1993). The more specialized (purified) the sciences become, the more they overflow their boundaries and get tangled up with each other and the world (hybridized). For example, the disciplined quest for a chemical to mix with gasoline to reduce engine knocking landed on lead, and then lead landed in the ocean and in the bodies of children and in Congress and courtrooms. In other words, the laboratory techniques that were at first isolated, later become entangled with ethics, values, and justice. The more powerful the technology, the more philosophical questions it will raise. This is why it is particularly important for philosophers of technology to get involved with actual cases in the real world and not be left on the sidelines discussing things only with their disciplinary peers. Finding ways to effectively communicate with engineers or Congress or consumers should be considered *the real philosophy*.

There is also a more prudential or self-serving reason for philosophy to break the radical monopoly of disciplinarity. Since the end of the Cold War, the university has increasingly found itself seeped in the politics of accountability or valorization. Philosophers and other academics are being asked to explain their value to society. The defining feature of this new politics is the quest for an explanation of impact that does not rely on serendipity. To "give an account" means to spell out how impacts happen, not to just appeal to some ill-defined, indirect happenstance. Politicians and parents (who often pay the skyrocketing costs of higher education) demand to know: what is the value of all this? The question is particularly important for philosophers and other humanists whose scholarship often goes uncited and even unread. Why should tenured philosophers get paid higher salaries than lecturers or adjuncts to teach fewer classes to produce articles and books that no one cares about?

That is the most important political question facing philosophers of technology. It is a question about relevance and impact. Unless philosophers of technology can convince parents, administrators, and policymakers that their research has value, they will continue to see a contraction of jobs that come with the time and money to do research. Tenure-track jobs have long been declining as academia moves toward a contingent labor force. The "teaching gig" is replacing the research job that used to come with a living wage, benefits, and security (Bousquet 2008).

But the problem of relevance or impact is not just a prudential matter of job security for professional philosophers of technology. It is also a scandal for philosophy understood in its perennial sense as the love of wisdom. Of course philosophers have almost always been marginal creatures operating on the edges of cultures dominated by religion, power, money, or entertainment. As a result, they often had to develop an esoteric message for one audience and an exoteric message for others. But that's just the point. They took rhetoric—questions of audience and framing—seriously. They did not, as is the case with disciplinary philosophy, sing in the same key to the same audience all the time. Philosophers of technology might in particular think of Karl Marx, who wrote in different registers for different audiences as journalist, political agitator, and scholar. The scandal is to isolate just one kind of writing or activity, purify it, and christen that as "real philosophy." This is especially true for philosophers of technology, given that we live in a high-tech world and the headlines daily present them with opportunities to get involved in real-world issues. Perhaps other kinds of philosophy (e.g., analytic metaphysics) are better suited to a professional remove from the hurly burly of daily life. But the philosophy of technology is all around us.

Further, disciplinarity has a way of turning philosophy (like the specialized sciences) into a technical enterprise. The ultimate ends and larger view are forgotten or obscured by a hyper-active race to produce the next knowledge unit. The imperative of specialization drives out consideration of the whole. An unspoken axiology valorizes cleverness and productivity above everything else. Lost is the once common-sense understanding that philosophers are seeking the good life or that philosophy might be a technique of self-improvement. How antiquated it seems to think that philosophers (despite their shortcomings) should strive to be model human beings and citizens. The *telos* or goal of disciplinary philosophy is to be smart, not good (Frodeman and Briggle 2016a).

Finally, consider how disciplinarity gives a rather un-philosophical (that is, unreflective or thoughtless) answer to questions about the philosopher's social responsibilities. They are held accountable only to their peers for producing knowledge units. This means that they are obligated to think of something deemed sufficiently novel and clever. Once their knowledge unit is deposited in the peer-reviewed reservoir, their work is done. Yet this is a rather impoverished account of responsibilities. Does the philosopher not have any wider obligations to society?

To use Winner's terminology, the discipline has an unavoidable "political cast" that is inward-looking. The specialized knowledge units produced are not designed to be useful *in a direct way* to anyone but fellow specialists. After all, we are talking about a purified region of discourse to which not just anyone can contribute. Knowledge contributions must first be certified by the expert gate-keepers, so anyone evaluated by this system must take these gatekeepers as their primary audience.

Though much good thinking can result from this, the system lacks a governor and so it eventually devolves into petty academic politics sometimes known as the genius contest. The result is a growing reservoir of peer-reviewed literature that may well be *about* real-world issues but that does not *speak to* or *influence* those issues in any practical sense. My own work on the windfarm is a case in point. That article is a line on my professional CV, but no one involved in the actual policy debate has ever referenced it.

In sum, the philosophy of technology needs elements both of purity and hybridity. The discipline provides a technique, a theory, and a dominant politics of purity. What is needed is a counter-balancing with techniques, theories, and politics of hybridity. In the last two sections, I make a start in that direction first by suggesting a taxonomy for alternative philosophies/technologies and then by proposing an agenda for political/philosophical reforms.

5. Alternative Technologies of Philosophy: A Bestiary

Imagine a catalog of alternative modes of transportation that fall outside of the radical monopoly of automobiles: walking, bicycling, roller skating, pogo-sticking, etc. That's my goal in this section: to theorize and categorize alternative philosophies of technology. Though this is better described as alternative technologies of philosophy, because what I want to foreground is not what people (conditioned by disciplinary ways of thinking) typically think of as "philosophy." I am not interested in the usual kinds of categories like ancient and modern philosophy or analytic and continental philosophy. Rather, I am interested in the *techniques* of philosophy—the practices, institutions, rhetorics, (re) production processes, and evaluation instruments—and how these techniques give rise to different politics or forms of life.

A word about the political focus of this taxonomy: these alternative modes of philosophy of technology are not just different ways to communicate. To varying degrees, they restructure relations of power and the very identity of the philosophy of technology. The field philosopher, for example, does not act as the expert delivering authoritative knowledge. Rather, they act as a participant in a case study. Their power derives far more from "street cred" than traditional modes of academic legitimacy. That is, they will be successful to the extent that they can earn the trust of those they wish to help. Further, what counts as "successful" or "excellent" work is determined in large part by those collaborators by their own standards rather than by disciplinary peers and the standards of the academy.

This taxonomy, then, presents variations on politics understood as "forms of life." This pertains to practices, norms, and far more. The reader is encouraged to imagine the different forms of life that would attend to each of these ways of being a philosopher of technology.

Consider this taxonomy:

This is a map of the different techniques and politics of the philosophy of technology. I will walk through it after a few preliminary remarks. First, this is neither a comprehensive nor a clean taxonomy—we could easily multiply and blend the boxes. All of these modes of philosophizing can complement each other, and any given philosopher can engage in any of them. So the distinctions do not indicate exclusions, but rather

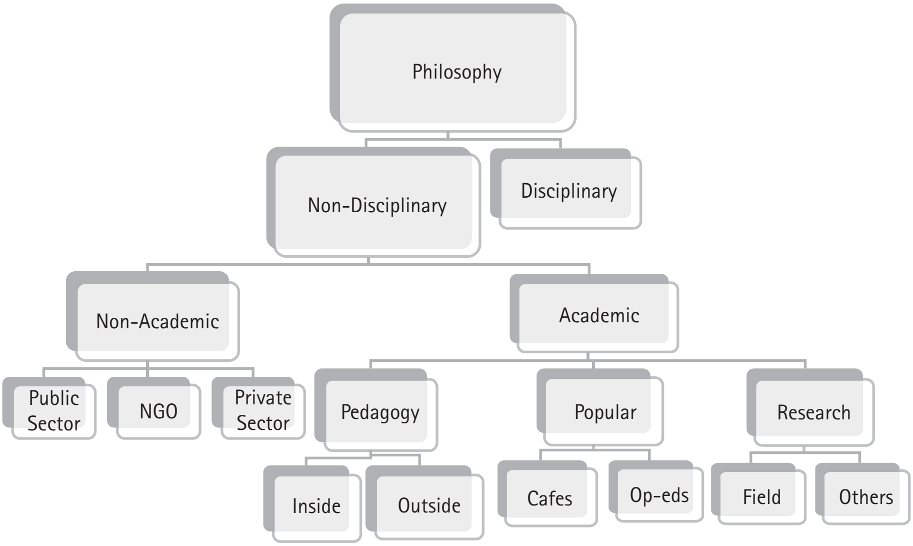

FIGURE 10.1: A Taxonomy of Philosophies.

Source: Author's own.

different techniques with their different or forms of life. Second, I am interested in the non-disciplinary techniques or practices of philosophy so I focus on those—one can imagine numerous branches under the box labeled "disciplinary" (e.g., analytic, modern, Kantian, philosophy of mind, etc.).

Third, although the radical monopoly of disciplinarity remains, non-disciplinary approaches to philosophy are proliferating. Philosophy of technology occurs with many practitioners and publics—engineers, policymakers, citizens, parents, consumers, industrialists, business executives, etc. Yet there has been little work to theorize or make sense of this proliferation. I intend this taxonomy as a preliminary attempt to lay out the important distinctions and relationships (see also Brister and Frodeman 2020). Finally, many philosophical schools, including Aristotelian practical wisdom, American pragmatism, and Marxism, provide historical (pre-disciplinary) examples of and conceptual foundations for non-disciplinary philosophical techniques. The important distinction is whether these schools are *put into practice* with non-philosophers or whether they are examined within disciplinary venues (i.e., as fodder for debate in a specialized journal or book). To put it crudely: is one practicing philosophy like Jane Addams or is one talking to other philosophers about Jane Addams?

The first-level distinction is between disciplinary and non-disciplinary philosophy. Non-disciplinary philosophy breaks in one way or another with the techniques of disciplinarity discussed earlier. This is principally about *with or for whom* one works: disciplinary peers or others. And again this is about politics, because these different audiences will serve as the gatekeepers for defining the work that is needed.

The term "applied philosophy" is often used to cover what I intend with "non-disciplinary philosophy." This leads to confusion, however, because a great deal of applied philosophy is disciplinary. Although this work is *about* a real-world issue (e.g., the ethics of a technology) it is written *for* consumption by fellow philosophers in the pages of specialized applied philosophy journals or at academic conferences. My windfarm paper is a good example of applied philosophy. In other words, it has succumbed to disciplinary capture and reductionism. For this reason, I favor "non-disciplinary" as a way to make the point about the *mode* (or politics) of the philosophical practice. Some—but certainly not all—applied philosophers do non-disciplinary work at least some of the time. Other terms are sometimes used to mean what I have in mind with non-disciplinary philosophy. These include public, engaged, and practical philosophy.

Further, it is worth noting that with partial exceptions (e.g., in biomedical ethics and conservation biology), applied philosophers have remained marginal to science, technology, and policy developments. This fact is known within the applied philosophy community: scholars have long lamented the failure of applied philosophy to live up to its aspirations of practical relevance (e.g., Amy 1984; Stone 2003; Lee 2008; Lachs 2009; Manson 2009; Hale 2011; Heal 2012; Wittkower, Selinger, and Rush 2013; Frodeman and Briggle 2015).

The next level down on the taxonomy is meant to capture the institutional home of the non-disciplinary philosopher. By "non-academic" I mean what Robert Frodeman

(2010) called "philosopher bureaucrats." These are philosophers employed outside of the academy (with a government agency, a non-governmental organization, or in the private sector) who help their organization think through the philosophical dimensions of their work. Examples include Damon Horowitz (former "in-house philosopher" for Google), Rene von Schomberg (STS specialist at the European Commission), and Johnny Hartz Søraker (Trust and Safety Policy at Google). In an example of a blended appointment, in 2019 Shannon Vallor remained in her academic position at Santa Clara University while also employed half-time as an AI Ethicist and Visiting Researcher for Google.

Non-disciplinary academic philosophers engage different audiences with different kinds of techniques. They can be housed in a philosophy department, elsewhere on campus, or both. Kathryn Plaisance, for example, holds a cross appointment in the Departments of Philosophy and Knowledge Integration at the University of Waterloo. Paul Thompson at Michigan State University is institutionally located in philosophy and two other departments that focus on sustainability and agriculture. Non-disciplinary academic philosophers could also rotate between departments, perhaps housed in philosophy but seconded for a year or more in an engineering department. Further, they could serve as in-house philosophers for Deans or Provosts in order to help them think about the roles and future of technology in the university.

The next level down highlights three main kinds of non-disciplinary academic philosophy, starting with pedagogical practices. Of course, most teaching is addressed to non-professional philosophers (i.e., students). Non-disciplinary pedagogy, however, goes further in at least two kinds of ways. What I call "inside" pedagogical practices are those that pertain to philosophy classes. They can incorporate service learning opportunities into the curriculum or bring non-philosophers into the classroom in various ways. This can also include explicitly problematizing the ways in which the class is typically taught by showing students how standard approaches might be captured by disciplinary concerns at the expense of forming meaningful relationships with the students' lived experiences. For example, philosophers of technology might give a traditional metaphysics course a twist by grounding it in questions about our increasingly cyborg existence. Beginning from students' experiences (say with their cell phones) rather than from academic literature is a simple way to de-discipline education.

"Outside" pedagogical techniques reach out to non-traditional students in non-traditional university classroom settings. In public philosophy circles, for example, there are thriving communities that do philosophy with prisoners and philosophy with children. I also include in this category what are often called "service" classes across the university that philosophy departments might offer. For the philosophy of technology, the best opportunities here have historically been teaching ethics to engineers, though technology is creating new venues such as the more recent rise of programs in big data or data analytics, which increasingly have ethics components. This teaching can occur with engineering students or with engineering professionals. For one example of the latter, Chinese philosophers of technology have for the past decade or more held ongoing workshops with high-level government engineers.

A second kind of non-disciplinary academic philosophy is "popular." Popular philosophers share their knowledge with wider audiences via blogs, videos, philosophical cafes, op-eds, and more. It can be a YouTube lecture on Hume or a documentary on the ethical dimensions of factory farms, etc. I would also include *Philosophy Now* and similar magazines in this category. And I put the "public intellectual" in this category, including contemporary philosophers such as Jürgen Habermas and Martha Nussbaum. These philosophers often contribute to contemporary popular discussions about important questions of technology and society in forums that are popular and accessible (e.g., newspapers or online magazines).

Public intellectuals are distinguished from other popular philosophers primarily by their platform or "brand recognition," which raises important questions about the politics of philosophy such as how one acquires a platform and what responsibilities and pitfalls come with it. As more philosophers of technology turn toward non-disciplinary practices, a crucial part of their training and professional lives will hinge on managing their public presence. Parts of this will be relatively simple, but other parts will raise philosophical problems in their own right. For example, in seeking to impact real-world social issues some philosophers of technology might be tempted to push controversial or even outlandish ideas. This can be an effective way to garner social media buzz, which could in turn be offered as evidence for one's impact. But there are obvious problems here, especially about advancing one's own career at the expense of causing chaos or even harm.

The third kind of non-disciplinary academic philosophy (and the main focus of this chapter) is "research," which might also be a good term for much of what non-academic philosophers do as well. As discussed previously, the disciplinary model of research entails the production of knowledge units filtered through peer-review and deposited in a reservoir of specialized knowledge. If this knowledge is to play a part in a real-time, real-world issue pertaining to technology, it will be through some intermediaries in an indirect and passive process. It is largely left up to the politics of serendipity. Non-disciplinary research seeks more self-conscious and direct impacts, which means that it takes non-philosophers (engineers, policymakers, public stakeholders, etc.) as the main audience, collaborators, or extended peer community. It also requires being evaluated by alternative standards or metrics. Non-disciplinary research is similar to "use-inspired basic research" (Stokes 1997) or the "scholarship of engagement" (Boyer 1996). It requires getting involved with an issue, gaining access to contribute, and earning the trust to be taken seriously (see Briggle 2020 for a detailed discussion of this).

Mitcham (1994b) proposed a research model for engineers called "curiosity *plus respicere*" or a "duty to take more into account." Non-disciplinary philosophical research is similar: it entails consideration not just of one's scholarly contributions but also one's "broader impacts" (from the US National Science Foundation) or one's "pathways to impact" (from the UK Research Councils). In contrast to the libertarian politics of the disciplinary technique, this mode of practice begins from a sense of situatedness and wider obligations. In a thin version, this could mean a kind of marketing plan to call attention to peer-reviewed publications once they are in print. But in richer versions, the expanded audience base would alter the very conduct of research—changing the practices, standards of rigour, the language, and the media used. In other words, this kind of research entails new techniques and new forms of life. Philosophical thinking grows beyond *what* ideas to ponder to include *how* to present ideas and to *whom*.

A philosopher might do "idea interventions" (rather than knowledge production) in a novel, a short video, a policy white paper, a community talk, an act of civil disobedience, etc. The activities or "form of life" would shift from reading articles and books to attending town hall meetings with activists, for example, or walking the production line with systems engineers. Knowledge or insights might come about far more through oral form in conversations in the midst of a collaboration than in a well-wrought written argument after ponderous thought. In other words, the knowledge is produced in the context of its use rather than in a self-contained knowledge unit such as a book or article.

There are many kinds of non-disciplinary research, including what Thompson calls "occasional philosophy" and Michael O'Rourke's "Toolbox Project" that facilitates philosophical dialogue within interdisciplinary research teams (see Brister and Frodeman 2020). Various forms of technology assessment work by philosophers also fits this category. For example, several Dutch philosophers of technology (e.g., Marianne Boenink and Tsjalling Swierstra) work with the Rathenau Institute to directly engage with technological developments. More generally, the Dutch 4TU Centre for Ethics and Technology includes in its research statement a commitment to contributing to "better practices" in engineering and technology policy. In Denmark, the Humanomics research project led by the philosopher David Budtz Pedersen maps the various pathways by which philosophy influences society in ways explicitly geared toward the needs of national and international policymakers. In the United States, Erik Fisher has developed methods for embedding philosophers and other humanists in science and engineering labs to seek "mid-stream modulation" of research projects (Fisher and Mahajan 2010).

In an interview with Mitcham, the Chinese philosopher of technology Yuan Deyu reports that prior to the 1980s, non-disciplinary research was common in China: "Chinese philosophers of technology originally began with real-world experiences. They generally attempted to learn technology first, then to analyze and philosophize about it" (Mitcham et al. 2018, 288). But under the influence of Western approaches, Chinese philosophers of technology switched to a disciplinary technique and "tended to reflect on and criticize technology based on some already existing philosophical system." A turn back toward non-disciplinary practices is occurring, however, among younger generations of Chinese philosophers of technology. For example, Yuan's student Yin Wenjuan at Northeastern University practices the philosophy of *gong cheng* (not quite translatable as "engineering") in direct partnership with engineers to help them think through the philosophical dimensions of their activities, language, and habits of thought.

With Frodeman, Britt Holbrook, Evelyn Brister, and others, my focus has been on "field philosophy" (Briggle 2015). Field philosophy takes inspiration from field science, as opposed to laboratory science, so the basic idea is to work with situations as they take their contested and messy shapes in the world rather than as they appear once they have been purified and reduced in the laboratory. The essential aspects of field philosophy follow:

- It involves case-based research at the project level: as a practice, it must consist of more than writing from the proverbial (or actual) armchair.
- It begins with the interests and framing of a non-philosophic audience rather than with the categories and interests of philosophers.
- The knowledge, insights, questions, or ideas it produces are done in the context of their use by non-philosophers.
- Its notion of rigor is contextual, sensitive to the demands of time, interest, and money.
- It prioritizes non-disciplinary standards for evaluating success (such as impacts on policy or fostering more thoughtful public debates).

Field philosophers can lead many forms of life. They will spend time at city hall, in the lab, on the farm, at the factory . . . and that is where the philosophy will happen. These are not resources to be harvested for raw materials and taken back to the ivory tower where the "real" philosophical work can be done. The philosophizing is in the interacting.

In all these kinds of non-disciplinary research, philosophers can offer much of value. They can help to identify, clarify, and critique conceptual and normative dimensions of the issue at hand. They can also muddle areas that are falsely clarified. Philosophers are good at challenging claims to expertise and authority, uncovering hidden value judgments and assumptions, recognizing and critiquing various arguments and framing devices, offering creative alternatives, and posing fundamental questions that are often overlooked.

As noted, many philosophers of technology already live these forms of life and make these kinds of contributions. Yet they have done so largely in spite of (not because of) their training and institutional incentives. And few have turned philosophy upon itself to theorize these new forms of life and take the political action to make them mainstream rather than marginal. The public philosopher Linda Martín Alcoff said that to do engaged philosophy is to "walk a fine line between responsiveness to community needs and employment survival, pushing the boundaries of academic respectability even while trying to establish . . . credentials in traditional ways" (Alcoff 2002, 522). But why not make responsiveness to community needs essential to employment? That should set the agenda for a reform of the philosophy of technology.

6. Conclusion: An Agenda for Reform

There is an argument to be made that philosophy is, by nature, not capable of being practically relevant. One could draw ammunition for this argument from Plato, Thomas

More, Hegel, and others (see Lee 2008). I grant the perennial tensions between philosophy and the *polis*, but philosophy can be—and, in fact, has been—of practical relevance to debates and practices involving technology. Further success is thwarted more by the technique of the discipline (an historical aberration) than by the essential features of philosophy.

Any political reform of technology, including the technologies of philosophy, calls for a host of strategic and tactical decisions. I think the goal should be the abolition of the radical monopoly of the discipline. In other words, the reform agenda should focus on ways to legitimize and incentivize non-disciplinary approaches to the philosophy of technology. In this, I agree with Dylan Wittkower, Evan Selinger, and Lucinda Rush (2013).

I will conclude with nine ideas that constitute my philosophical/political reform agenda. Philosophers of technology should

1. Develop and share standards for hiring, promotion, and tenure that reward non-disciplinary philosophy (see Lachs 2009). Some departments already have unorthodox standards in place, but most do not. Indeed, a survey of humanities departments in the United States found that only 5 percent considered "public humanities" either essential or very important for tenure and promotion valuation (see Frodeman and Briggle 2016a).
2. Develop and share alternative metrics (altmetrics) for research evaluation capable of capturing the impacts of non-disciplinary philosophy (see Wilsdon et al. 2016). Departments might consider keeping an updated list of their impacts on their websites and more generally foster ways to value activities and products that are not standard, peer-reviewed publications. This can also be done through novel publishing practices. For example, Wittkower's new *Journal of Sociotechnical Critique* employs alternative techniques to validate engaged scholarship.
3. Create awards for engaged philosophy to foster a "challenger axiology" to the default, disciplinary valuation of so-called "real" philosophy. So far, few such awards exist. The APA Public Philosophy Award should be significantly bolstered. The Public Philosophy Network is likely to step into this area over the coming years.
4. Develop pedagogical tools, workshops, and curricula for training next-generation scholars in non-disciplinary philosophical practices. Non-disciplinary techniques are too often created as one-off re-inventions. The philosophy of technology community needs to think about what counts as smart practices for different kinds of politics and how those practices can be conveyed to the next generation (see Brister and Frodeman 2020).
5. Create internship programs to place graduate students in the public and private sector and to foster relationships that can generate non-academic careers. Few companies or government agencies think of their issues as having philosophical dimensions, and so they rarely advertise jobs for philosophers. Yet they might take free help from interns who might in turn convince them of the value of philosophical engagement. We should take an experimental approach to this to figure out

how to do the philosophy of technology outside of academia in all sorts of venues. Many experiments will fail, but the important thing is to learn from them.

6. Create novel institutional arrangements across campus to second philosophers in other departments temporarily to aid with their research projects. One ready-made mechanism for this is the Broader Impacts criterion at the National Science Foundation. Philosophers of technology could serve as in-house consultants and partners to help research teams articulate, track, and manage the broader social impacts of their work.
7. Develop new theoretical accounts of rigor—a different kind of hard—that can be used to describe and teach the skills involved in non-disciplinary work. So much of the "soft power" of the discipline comes from an assumption that it represents the only "real" kind of hard work properly understood as philosophical excellence. This needs to be exposed as a sham by developing rich accounts of the skills involved in doing engaged philosophy with excellence and care.
8. Challenge the libertarian politics of disciplinarity by conducting joint strategic planning exercises to identify strengths and surrounding community needs. Philosophers of technology should think of themselves as handmaids to their community. Their prime directive should be to help people think through the implications of decisions about technologies. Being a servant does not mean being servile or "selling-out" to say whatever a client wants to hear. Philosophy must always retain a critical edge and a gad-fly sensibility. But too often that has devolved into utter abdication of the public sphere and irrelevance. It is time to rectify this imbalance.
9. Develop and share codes of ethics to guide philosophers in the conduct of non-disciplinary research. As already noted, the ethics of disciplinary work is too restricted in considering obligations only to fellow experts. Non-disciplinary practices broaden the scope of obligations. They also muddy the picture: what ideals should guide engaged philosophers of technology and how should they handle the conflicts and ambiguities that inevitably arise when getting one's hands dirty? And how do we account for the new political dynamics introduced when one expands the "peer" base beyond the discipline?

This is not a comprehensive agenda, but hopefully it is enough to suggest the richness and diversity of political *and* philosophical issues that arise when challenging the orthodox technology of philosophy.

References

Alcoff, Linda Martín. 2002. "Does the Public Intellectual Have Intellectual Integrity?" *Metaphilosophy*, 33, no. 5, 521–534.

Amy, Douglas J. 1984. "Why Policy Analysis and Ethics Are Incompatible," *Journal of Policy Analysis and Management*, 3, no. 4, 573–592.

Bordogna, Francesca. 2008. *William James at the Boundaries: Philosophy, Science, and the Geography of Knowledge*. Chicago: University of Chicago Press.

Borgmann, Albert. 1995. "Does Philosophy Matter?" *Technology in Society*, 17, no. 3, 295–309.

Bousquet, Marc. 2008. *How the University Works: Higher Education and the Low-Wage Nation*. New York: New York University Press.

Boyer, Ernest L. 1996. "The Scholarship of Engagement." *Journal of Higher Education Outreach and Engagement*, 1 (Jan.), 11–20.

Briggle, Adam. 2015. *A Field Philosopher's Guide to Fracking*. New York: W.W. Norton.

Briggle, Adam. 2016. "The Policy Turn in the Philosophy of Technology." In *Philosophy of Technology after the Empirical Turn*, edited by Maarten Franssen, Pieter E. Vermaas, and Anthonie W.M. Meijers,167–176. Cham, Switzerland: Springer.

Briggle, Adam. 2020. "Learning from a Fracking Fracas," In *A Guide to Field Philosophy: Case Studies and Practical Strategies*, edited by Evelyn Brister and Robert Frodeman, 50–69. New York: Routledge.

Brister, Evelyn, and Robert Frodeman. 2020. *A Guide to Field Philosophy: How to Use Philosophy to Change the World*. New York: Routledge.

Bush, Vannevar. 1945. "Science—the Endless Frontier." Washington, D.C.: US Government Printing Office: http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm.

Callahan, Daniel. 1973. "Bioethics as a Discipline." *Hastings Center Studies*, 1, 66–73.

Callicott, Baird. 1999. "Environmental Philosophy *Is* Environmental Activism: The Most Radical and Effective Kind." In *Beyond the Land Ethic: More Essays in Environmental Philosophy*, 27–44. Albany, NY: State University of New York Press.

Carnap, Rudolph. 1963. *The Philosophy of Rudolf Carnap*, edited by P. A. Schilp. La Salle, IL: Open Court.

Changshu, Chen. 2016. *An Introduction to Philosophy of Technology*. Translated by Chen Fan, Ma Ming, and Howard Giskin. Beijing: Science Press.

Fisher, Erik, and Roop L. Mahajan. 2010. "Embedding the Humanities in Engineering: Art, Dialogue, and a Laboratory." In *Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration*, edited by Michael Gorman, 209–231. Cambridge, MA: MIT Press.

Frodeman, Robert. 2010. "Experiments in Field Philosophy." *New York Times*, Nov. 23.

Frodeman, Robert, and Adam Briggle. 2015. "Socrates Untenured." *Inside Higher Ed*, January 13.

Frodeman, Robert, and Adam Briggle. 2016a. *Socrates Tenured: The Institutions of 21st Century Philosophy*. Lanham, MD: Rowman and Littlefield.

Frodeman, Robert, and Adam Briggle. 2016b. "When Philosophy Lost Its Way." *New York Times*, January 11.

Frodeman, Robert, Julie Thompson Klein, and Roberto Carlos Dos Santos Pacheco, eds. 2015. *The Oxford Handbook of Interdisciplinarity*. 2nd ed. New York: Oxford University Press.

Hale, Ben. 2011. "The Methods of Applied Philosophy and the Tools of the Policy Sciences." *International Journal of Applied Philosophy*, 25, no. 2, 215–232.

Heal, Jane. 2012. "Philosophy and Its Pitfalls." In *The Pursuit of Philosophy: Some Cambridge Perspectives*, edited by Alexis Papazoglou, 37–43. West Sussex, UK: Wiley-Blackwell.

Heidegger, Martin. 1977. *The Question Concerning Technology and Other Essays*. New York: Harper and Row.

Illich, Ivan. 1973. *Tools for Conviviality*. New York: Harper & Row.

Kroes, Peter, and Anthonie Meijers, eds. 2000. *The Empirical Turn in the Philosophy of Technology*. Amsterdam: Elsevier.

Krohn, Wolfgang. 2010. "Interdisciplinary Cases and Disciplinary Knowledge." In *The Oxford Handbook of Interdisciplinarity*, edited by Robert Frodeman, Julie Thompson Klein, and Carl Mitcham, 32–49. New York: Oxford University Press.

Lachs, John. 2009. "Can Philosophy Still Produce Public Intellectuals?" *Philosophy Now*, 75, 24–27.

Latour, Bruno. 1993. *We Have Never Been Modern*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 2007. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Lee, Steven. 2008. "Is Public Philosophy Possible?" *International Journal of Applied Philosophy*, 22, no. 1, 13–18.

Manson, Neil. 2009. "Epistemic Inertia and Epistemic Isolationism: A Response to Buchanan." *Journal of Applied Philosophy*, 26, no. 3, 292–298.

Mitcham, Carl. 1994a. *Thinking through Technology: The Path between Engineering and Philosophy*. Chicago: University of Chicago Press.

Mitcham, Carl. 1994b. "Engineering Design Research and Social Responsibility." In *Ethics of Scientific Research*, edited by Kristin Shrader-Frechette, 153–155. Lanham, MD: Rowman and Littlefield.

Mitcham, Carl, Bocong Li, Byron Newberry, and Baichun Zhang, eds. 2018. *Philosophy of Engineering, East and West*. Cham, Switzerland: Springer.

Polanyi, Michael. 1962. "The Republic of Science." *Minerva*, 1, no. 1, 54–74.

Stokes, Donald. 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, D.C.: Brookings Institution Press.

Stone, Christopher. 2003. "Do Morals Matter? The Influence of Ethics in Courts and Congress in Shaping U.S. Environmental Policies." *Environmental Law and Policy Journal*, 27, no. 1, 13–35.

Turner, Stephen. 2000. "What Are Disciplines? And How is Interdisciplinarity Different?" In *Practising Interdisciplinarity*, edited by Peter Weingart and Nico Stehr, 46–65. Toronto: University of Toronto Press.

Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. Chicago: University of Chicago Press.

Veysey, Laurence 1979. "The Plural Organized Worlds of the Humanities." In *The Organization of Knowledge in Modern America, 1860–1920,* edited by Alexandra Oleson and John Voss, 51–106. Baltimore, MD: Johns Hopkins University Press.

Wilsdon, James, Judit Bar-Ilan, Robert Frodeman, Elizabeth Lex, Isabella Peters, and Paul Wouters. 2016. "Next-generation Metrics: Responsible Metrics and Evaluation for Open Science." Report of the European Commission Expert Group on Altmetrics. Brussels: European Commission.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus*, 9 no. 1, 121–136.

Winner, Langdon. 2014. "Technologies as Forms of Life." In *Ethics and Emerging Technologies*, edited by R. L. Sandler, 48–60. London: Palgrave Macmillan.

Wittkower, D. E., Evan Selinger, and Lucinda Rush. 2013. "Public Philosophy of Technology: Motivations, Barriers, and Reforms." *Techne: Research in Philosophy and Technology*, 17, no. 2, 179–200.

CHAPTER 11

POSTCOLONIALISM AND TECHNOLOGIES OF IDENTIFICATION

ALISON ADAM

1. Introduction

Philosophy of science and technology is interwoven with the history of science and technology, where, at a minimum, history provides philosophy with its examples. Important "turns" in the history and philosophy of science and technology have influenced the ways in which technology and science are understood by philosophers and historians. From the 1960s onwards, history and philosophy of science and technology experienced a "social turn" where scientific and technical arts were increasingly seen as social products rather than the results of an "internalist" scientific method. According to the internalist approach the scientific method resists the taint of external influence such as so-called "social factors," proceeding instead in terms of its own internal logic which produces scientific knowledge, which is then translated into technology in the form of applied science. This older view of the creation of scientific knowledge, what might be termed the "modernist" approach, held that scientific knowledge traveled when it was true (Shapin 1998, 7). For instance, one of the founding fathers of history and philosophy of science, George Basalla, developed a much criticized diffusionist model of the spread of western scientific knowledge (Anderson 2002, 648). Basalla's model held that western science spread from center to periphery, where colonial expeditions provided examples for western science and where the development of colonial institutions were dependent on importing western scientific and technological knowledge. Although independent scientific institutions might develop in colonial settings, Basalla regarded the flow of true scientific knowledge as basically one-way, diffusing from western settings to peripheral colonial settings which were ready to receive it (Anderson 2002, 648; Basalla 1967).

The main issue with the traditional philosophy of science and technology characterized there, is that it leaves unchallenged a hegemonic view of the creation and movement of scientific and technical knowledge. The center and periphery model assumes the creation of scientific and technical knowledge in Europe which then moves into a colonial setting, illustrating that true knowledge will naturally travel to the periphery like water flowing down a hill under the force of gravity. In the 1980s, Basalla's model was heavily criticized within the wider interdisciplinary field of science and technology studies (STS), which incorporates philosophical, historical, and sociological approaches. Notably, Roy MacLeod (1982) disputed the linearity of the diffusionist approach and its avoidance of political factors, arguing for the study of reciprocity and the complex nature of contact cultures, the conceptual space where people of different cultures, histories and geographies come together. Palladino and Worboys (1993) argued for an understanding of the ways in which western methods were adapted into existing traditions of knowledge rather than just being passively accepted. Although not addressing the question of the influence of non-western cultures in the development of science, Thomas Kuhn's (1962) *The Structure of Scientific Revolutions* was highly influential in offering a challenge to the "logic of scientific progress," which helped to spark the social constructionist push in philosophy and science and technology studies.

For some, the apparent attack on the logic of scientific progress implied by increasing interest in understanding science and technology as social and cultural products has been hard to accept. Nevertheless, over the last forty and more years the decentering of scientific and technological progress has been firmly embedded in the philosophy of science and technology. Yet the social turn in science and technology studies has not turned far enough, as it leaves the voice of the western white man as the knowing subject firmly, if often implicitly, in charge of knowledge production. It is difficult to escape the view that the superiority of this knowing subject's view of science and technology is still tacitly accepted. Hence, from at least the 1970s onwards, feminist science and technology studies (particularly feminist epistemology) and postcolonial approaches to science and technology have continued to question the assumed superiority of the white western masculine view, the position that Basalla and modernist models implicitly assumed (Harding 1998, 2008, 2011).

However, even if models of the spread of scientific and technological knowledge have benefitted from a "social turn" that challenged the invisible hegemony of the white western male in the making of such knowledge, on its own a social turn cannot adequately explain the spread of scientific and technological knowledge. Arguably a "geographical turn" is necessary to achieve such an understanding. Fortunately, STS is now well advanced in the process of taking a "geographical turn" which involves developing new postcolonial understandings of the production of local knowledges and the way in which such knowledge travels, at the same time acknowledging the political dimension of the movement of scientific and technological knowledge (Shapin 1998).

The postcolonial approach to STS has developed rapidly over the last few decades. The richness and multi-faceted nature of the burgeoning collection of STS studies adopting one or another variant of postcolonialism makes it impossible to review the complete range of literature here. Hence after a very brief introduction to postcolonialism, this chapter sets out the theoretical elements of postcolonialism in relation to STS which are relevant to technologies of identification. Arguing that the philosophy of science and technology is best illuminated by historical examples, I consider the work of Kapil Raj (2007) on the construction and circulation of knowledge in south Asia and Europe. I offer two examples as illustrations of postcolonial philosophy applied to historical examples of forensic epistemology; these center on the development of fingerprinting, and the translation and adaptation of a European book on criminalistics (nowadays known as forensics in North America and forensic science in the United Kingdom) in India at the turn of the nineteenth century and early twentieth century. Although Raj's case studies derive from an earlier period than my examples, there are many continuing similarities in terms of contact zones, trust relationships in the making of knowledge, and the information order.

It is clear that many of the negative aspects evident in historical case studies remain relevant to philosophy of technology's critique of contemporary technologies of identification. Indeed, the examples of fingerprinting and the development of criminalistics described here were arguably some of the first attempts to use identification technologies for mass control and surveillance of populations, for civil and criminal justice purposes. The story of their development shows how identification technologies can be used to demonize or criminalize certain ethnic groups. The colonial use of these technologies provides some of the earliest models of how an ethnic group can be suppressed by the use of such technology, a model which, for good or ill, still persists. A pertinent contemporary example is to be found in automatic face recognition software and how it is currently used in some parts of the world (Byler 2019), as I describe briefly in the conclusion.

2. Postcolonialism

Space permits only the briefest introduction to postcolonialism rather than providing a comprehensive description. The aim is to set the context for a more detailed discussion of postcolonial theory as it has been applied in the philosophy of science and technology. Postcolonial theory is a recent body of theory and a set of empirical studies which examine the legacy of colonialism and its impact, emphasizing control and an ongoing history of exploitation (McNeil 2005, 106–107). McNeil argues that postcolonialism is a somewhat ambiguous term referring both to the aftermath of oppressive colonial systems and new forms of exploitation. As a theoretical framework it provides the scaffolding to unpick assumptions of European science and technology as a "culture of no culture" (Traweek 1988) and to understand European science as "situated knowledge," (Harding 1998, 14) just as much as science and technology are understood anywhere else in the globe.

Edward Said's (1978) highly influential although criticized book, *Orientalism*, which disparaged western societies' tendency to exoticize representations of Middle Eastern cultures, was a significant source of postcolonial thought, particularly in the area of literary criticism. The feminist author Gayatri Chakravorty Spivak (1993) has been influential in moving the subject matter of postcolonialism beyond literary studies to philosophy. Her emphasis on subalterns or those colonial people who are excluded from power, especially subaltern women, and the ways in which subalterns are excluded from western philosophical theory continues to indicate a reorientation of "postcolonial theory away from literary canons towards alternative local knowledge production." (McNeil 2005, 107) Importantly, Spivak emphasizes local knowledges which can be made visible by giving a voice to those rendered invisible by colonialism (Anderson 2002, 646).

Warwick Anderson (2002, 645), in broad agreement with McNeil's (2005) view of postcolonialism also emphasizes the ambiguous nature of the postcolonial: "It has been taken to signify a time period (after the colonial); a location (where the colonial was); a critique of the legacy of colonialism; an ideological backing for newly created states; a demonstration of the complicity of Western knowledge with colonial projects; or an argument that colonial engagements can reveal the ambivalence, anxiety and instability deep within Western thought and practice." Three main themes are identified by Anderson: colonial critique, postcolonial theory, and historical anthropology. While postcolonial theory challenges the objectivity of western knowledge, nevertheless historians and anthropologists have criticized its reductiveness and tendency to universalize. These critics argue against potentially homogeneous categories of "colonial discourse," looking instead for more specific examples of agency and locations (Thomas 1994; Anderson 2002, 647).

3. Postcolonialism and Philosophy of Science and Technology

Acknowledging the wide-ranging yet ambiguous scope of postcolonialism, we now turn to consider how it may fruitfully be applied to STS, with particular reference to ways in which it might inform the "geographical turn" to understand the movement of scientific and technological knowledge through particular historical examples. Although there are signs that this is now changing, STS as a discipline has tended to be Eurocentric, perpetuating "the assumption that modern science is trans- or a-cultural" (Harding 1998, 14). Postcolonialism therefore becomes an important tool in the armory to challenge core-periphery diffusionist models and universalist conceptions of scientific and technological knowledge (Schiebinger 2005, 53).

A useful framing of postcolonial approaches to philosophy within STS starts with an interest in the exploration of "geographical sensibilities." Steven Shapin sees these as a development from "tensions in science studies between transcendentalist conceptions of truth and emerging localist perspectives on making, meaning and evaluation of scientific knowledge" (Shapin 1998, 5). Traditional philosophy of science's adherence to the universality of scientific knowledge, alluded to earlier, means that a "geographic approach" in STS has only begun to gain momentum in the last thirty years or so. The older concept of an inner logic to scientific knowledge has proved obdurate. A kind of half-way house was popular in STS for a while, where the production of scientific ideas was acknowledged as geographically contingent but the translation of scientific ideas to knowledge was not. A geographical approach to STS has been feasible from the 1980s onwards, as the sanctity of the inner logic of scientific knowledge had been definitively breached by then. This allows us to see that science and technology are local practices, which can nevertheless travel (Anderson 2002, 649). Such an approach has been influential in Actor-Network Theory as developed by John Law and others (Law and Hassard 1999). "Once knowledge can be analysed in terms of region, domain, implantation, displacement, transposition, one is able to capture the processes by which knowledge functions as a form of power and disseminates the effects of power" (Foucault 1980, 69 quoted in Shapin 1998, 6).

For Shapin (1998), writing some twenty years ago, the geographical turn in STS was something of an achievement. Still, at that time the turn had not gone far enough, as it is not just a question of uncovering how knowledge is made in specific locations but what kind of transactions occur between places for knowledge to spread. Even if we reject the older modernist view that knowledge which is true will travel because it is true, and that which is not true will not travel, we still have to explain why scientific and technological knowledge manages to travel so effectively.

Latour (1987) contends that it is a question of standardization. Once knowledge is standardized and institutionalized it may travel in an unmodified way—think of maps, thermometers, slide rules and the like. However, Shapin has criticized Latour's view of the transport of knowledge for not attending to the normative order, for being "all natural fact and no moral fact" (Shapin 1998, 7). *Trust* in the transport of scientific knowledge is an important aspect of its potential to travel, as is evident in the examples later in the chapter. Thinking of the origins of modern science, seventeenth century natural philosophers could obtain knowledge by observing a phenomenon directly or, more likely, by being told through the medium of an appropriately trusted person or that person's writings, even though they were not, themselves, direct witnesses. Those of an appropriately gentlemanly disposition were deemed trustworthy witnesses; hence those who had not seen something directly could achieve knowledge from those who had, and so on. For Shapin (1998, 8) the geographic sensibility in STS must address the question of who can be trusted, and thereby how distant knowledge may be brought close.

The abandoning of modernist notions of "true" knowledge in STS was part of the adoption of the *symmetry principle*, first in terms of the treatment of true and false belief and, later, in terms of the treatment of human and non-human actants in Actor-Network Theory (Law and Lin, 2017, 211). In accordance with Shapin's call for a geographical turn in STS, Law and Lin (2017, 222) argue that it is time for an extension of the principle of symmetry again, this time to a "postcolonial symmetry":

> In this the discipline would *explore the politics and analytics of treating non-western and STS terms of analysis symmetrically*. This means that STS would *stop automatically privileging the latter*. It would abandon what Warwick Anderson and Vincanne Adams aptly call the '"Marie Celeste" model of scientific travel' in which analytical terms (or laboratories or facts) travel silently and miraculously from metropolis to periphery . . . Instead in this postcolonial version of symmetry the traffic would be lively, two-way, and contested
>
> (Law and Lin 2017, 214).[1]

Anderson and Adams (2008) have explored the question of the mobility of knowledge, focusing on the useful concept of *technoscience*, which signals the entwined nature of science and technology, indeed the impossibility of separating the two concepts in many cases. This view of the interweaving of often diverse and heterogeneous scientific, technological, political and bureaucratic elements into "technoscience" (Adam 2016, 182) has become popular in STS. Anderson and Adams (2008, 184–185) emphasize the wide distribution of spaces where technoscientific knowledge is made—knowledge is not created just in the laboratory. They argue against authors who are determined to show that there was no colonial "contamination" or "entanglement" of technoscientific knowledge in the exact sciences, arguing instead that postcolonial scholars should "reveal the heterogeneity and messiness of technosciences" (Anderson and Adams 2008, 187).

Acknowledging messiness and heterogeneity, one of the most important aspects of postcolonial approaches to philosophy of science and technology is the analysis of ways in which ideas and artifacts are transported to or from the colonial locus to other places, including western settings, and how such movements of knowledge and artifacts are to be acknowledged and understood. A postcolonial approach can be used to question the implicit assumptions in earlier history and philosophy of science and technology, which emphasized and even valorized activities such as building up vast museum collections of tools and artifacts from other cultures brought back to the west.

As Mary Terrall (2011) notes, heroic tales of "quest and discovery" were a significant part of the story of the development of science and technology at least from the time of the Scientific Revolution. The metaphor of illuminating, uncovering, and displaying dark corners of the world was significant, yet the direction of the beam was always seen as emanating from Europe and was to be directed from there towards "exotic" cultures, rather than the other way around. Those who undertook voyages of discovery combined manly attributes of strength, courage, and endurance with intellectual brilliance, "and provided the raw material for secondary historical accounts of discovery and the progress of knowledge that grounded later versions of the grand narrative of scientific progress" (Terrall 2011, 85).

Collecting facts and artifacts from distant places was one aspect of colonial histories of science and technology, which have traditionally supplied philosophy of science and technology with their examples. However, instead of thinking about the removal of artifacts from their context to fit their dried husks into western taxonomic schemes, a more thoroughgoing postcolonial approach to the philosophy of science and technology eschews the suggestion that "civilization" was exported *to* far-flung lands, arguing instead that science and technology were often developed *in situ*, a matter of co-production in a colonial setting and then exported to the west (Raj 2007). The direction of travel was the opposite of that assumed by Basalla (1967) and the modernist tradition.

Taking the philosophy of technology as a focal point, in what ways can a postcolonial approach provide a "geographical turn" to develop a more nuanced understanding of the creating and movement of scientific and technological knowledge from east to west and *vice versa*, one which decenters western hegemony in the making of technoscientific knowledge? Historical examples are useful to put flesh on the bones of what might otherwise be construed as a somewhat abstract approach to postcolonialist philosophy of technology, or more properly, the philosophy of technoscience. The work of Kapil Raj (2007) on the circulation and construction of knowledge in British India in the eighteenth and nineteenth centuries provides an important analysis.

4. The Circulation of Knowledge

Kapil Raj, in his study of knowledge production in South Asia from 1650–1900, criticizes simple diffusionist models, emphasizing instead the co-production of knowledge in the intercultural "contact zone" of Europe-South Asia (Raj 2007, 11). Raj follows Mary Louise Pratt's (1992, 6–7) useful characterization of a contact zone as a meeting space for people of different cultures, histories and geographies, although almost always involving discrimination, struggle and oppression. Nevertheless, such contact zones can be places where scientific and technological knowledge is made, particularly knowledge which later comes to be regarded as western knowledge and where the hybrid nature of its construction is forgotten or ignored.

Raj's work joins the newer scholarship in history and philosophy of science and technology, which emphasizes detailed case studies of the ways in which scientific knowledge, technology, skilled practices, procedures and instrumentation were created in local, contingent, situated settings over attempts to construct grand narratives of the making and movement of knowledge. Such case studies demonstrate that the construction of scientific and technical knowledge depends less on an inner logic of scientific reasoning and more on the practical arts, in local settings of laboratories, libraries, hospitals, museums, botanical gardens and so on. But if such knowledge is not universal, how may we account for the movement of scientific knowledge, practices and instrumentation? "[T]hey disseminate only through complex processes of accommodation and negotiation, as contingent as those involved in their production" (Raj 2007, 9).

Raj (2007, 12–13) argues against using terms such as "colonial science" or "colonial knowledge," as these were terms that European colonizers used in relation to indigenous knowledge. Hence their use acknowledges the local nature of some knowledge, which is then not accorded the status of true scientific (i.e. western) knowledge. Instead, his aim is to promulgate "an alternative vision of the construction and spread of scientific knowledge through reciprocal, albeit asymmetric, processes of circulation and negotiation" (Raj 2007, 13). The making and circulation of scientific and technological knowledge in the colonial period took place in a much wider setting than scientific laboratories and museums. As my example will illustrate, religious mission organizations were important in the making and spread of technological knowledge. Overseas trade was also hugely important; for example, the agricultural trade in profitable plants. Organizations such as the East India Company were users and makers of scientific and technological knowledge. In nineteenth century South Asia a growing number of graduates from Scotland and North European universities began to occupy senior technical positions in trading companies; the "geographies of trade and knowledge thus largely overlapped" (Raj 2007, 17–19). Similarly qualified Northern Europeans continued to take up government scientific and technical posts at the end of the nineteenth century and beginning of the twentieth, as my examples will demonstrate. Raj emphasizes not just the historical contingency of the making of knowledge but also the *mutation* that movement engenders: "their transformations and reconfigurations in the course of their geographical and/or social displacements, that the focus on circulation helps bring to the fore" (Raj 2007, 21). We will see this illustrated in my case study on criminalistics, what we now call forensics or forensic science, where technological knowledge mutates when moving from west to east and back to the west.

A crucial part of the making of knowledge is the question of its certification, particularly trust and authority which points to the question of who is to be trusted. The nineteenth-century colonial Indian setting which Raj describes is of course far from the seventeenth-century English setting of Shapin's (1994; 1998) writing on scientific trust. Scientific trust has enormously increased in complexity since then, but the original concept recognizably persists. As Raj notes, seventeenth-century men of science did write manuals for those who traveled and collected to show them how the world should be viewed scientifically (Raj 2007, 103). Thinking about how this was achieved in the seventeenth century makes explicit some of the aspects of the involvement of indigenous people in the making of knowledge. Although some European travelers disguised the input of indigenous people, the question of how local, indigenous people were to be involved was always seen as important:

> even a cursory reading of these instructions makes it clear that almost all knowledge produced outside the well-defined precincts of the metropolitan laboratory

> implied the active participation of indigenous collaborators. And, while following the instructions and recipes of social promotion enabled British producers of these new knowledges simultaneously to acquire gentlemanly civility and settle authority claims, there remained the thorny question of the 'other' civility—hence credibility—of indigenous interlocutors on whose linguistic means and testimony much new knowledge and associated material practices depended. Certain instructions to travellers quite explicitly required the enlisting of autochthonous cooperation.
>
> (Raj 2007, 103)

Although more than century before my case study, Raj's discussion of the experiences of Sir William Jones, appointed as a junior (puisne) judge in the Supreme Court at Calcutta towards the end of the eighteenth century (Raj 2007, 119) has particular resonances, at least in part as it serves to underscore the relationship between legal and technoscientific knowledge that is central to scientific criminal identification technologies.[2] Importantly, Raj argues that the traditional view of historians, namely that trust was centered in the political and social superiority of the British rulers, masks the subtlety of trust relationships in the making of knowledge in colonial India. Trust and civility "underpinned the establishment of sustained relationships between rulers and ruled, without which colonial rule could not have been instituted let alone maintained for almost two centuries" (Raj 2007, 106). Raj argues that Jones was a pioneer of "the institutionalization of intercultural trust in the production of administration" (Raj 2007, 106). This was a model for the institutionalization of other kinds of administrative knowledge as part of the colonial information order. In an important sense the relationship between Britain and India progressed to being a collaboration over information including, as time went by, taxes, judicial administration, and education, where many already existing administrative structures had to be maintained. In his judicial role, Jones soon came to understand the reliance of the British on indigenous intermediaries in the administration of justice. At the end of the eighteenth century the British sought to stabilize legal and administrative codes in written form and for this undertaking they had to trust the testimony of indigenous experts (Raj 2007, 123–126).

Because of structural similarities between British and Indian society, Jones and his colleagues were able to collaborate with the appropriate levels of Indian society and invent a set of conventions to facilitate this collaboration. Hence, Raj argues against the traditional view of Indians as passive informants who had European beliefs imposed upon them, arguing instead for "an active, though asymmetrical, indigenous cooperation both in the *making* of new administrative knowledge . . . and in the moulding of British and Indian civilities in such a way as to render them commensurable" (Raj 2007, 138). As Raj demonstrates, within a colonial setting, the development of the information order with appropriate machinery of administration was well underway in the eighteenth century. The next step in the expansion of administrative control involves the development of appropriate identification technologies to support civilian government and criminal justice.

5. The Development of Fingerprinting Technology in Colonial India

India at the end of the nineteenth century and beginning of the twentieth century provided fertile ground for the development of technologies for the unique identification of individuals. The story of the most familiar of these, fingerprinting, is well-known and there are some excellent recent histories (Cole 2001; Sengoopta 2003). The perceived problem in colonial India was initially a civil problem, namely the reliable, unique identification of an individual for civil purposes such as contracts, pensions and other benefits. The problem was compounded by an environment where many were not literate, coupled with the British colonial authority's mistrust of indigenous, especially nomadic, people (Adam 2016, 106–107). It is notable that the civil applications of identification technologies were regarded as important as, if not more important than, criminal justice applications. Colonial administrator Sir William James Herschel (of the famous astronomical dynasty), working in Jungipoor and anxious about the potential repudiation of a contract at a later date, asked a local contractor to "sign" a contract with a handprint in 1858. Thereafter, he became fascinated by fingerprints, amassing a huge collection which he began to use in his role as a magistrate to cut down impersonation and repudiation of contracts, with some success.

When Scottish missionary Henry Faulds, working in Japan, wrote to Charles Darwin in 1880 on the possibility of using fingerprinting for criminal identification, Darwin passed Faulds's letter to his cousin, Francis Galton. A leading scientist of his day, Galton championed eugenics, was an expert in biometrics and was fascinated by hereditary genius, his own included, no doubt prompted by his links to the eminent Darwin-Wedgwood family. Galton corresponded enthusiastically with Sir William Herschel, who was of appropriate scientific and gentlemanly lineage, while ignoring the contribution of Faulds, an unknown provincial doctor. Such was the fame and prestige of Galton that his book on fingerprinting did much to publicize the possibilities inherent in using fingerprinting as a candidate for an identification technology (Galton 1892; Adam 2016, 108–109).

However, in order to be of any use for unique identification of possible offenders in a criminal justice system, an appropriate classificatory system had to be developed. One of the earliest effective systems was developed in a colonial setting designed for what were seen to be peculiarly colonial classificatory problems. British official Sir Edward Henry joined the Indian Civil Service in 1873 and became Inspector General of the Bengal Police in 1891. Henry introduced the Bengal police to anthropometry, the system of identification of individuals by recording a set of precise body measurements as developed by Alphonse Bertillon (Adam 2016, 109). Anthropologists were already using anthropometric measurements to measure purported differences between castes in India, but Henry's interest centered on systems to identify members of the so-called "criminal tribes" which were causing so much concern to the British administration.

The fingerprinting example involves an important contact zone for the circulation of local and western knowledge as British colonial interests and the initial development of fingerprinting by western scientists and administrators was brought into contact with the local knowledge of Indian technical officials to form a workable system. In developing a classificatory system for fingerprinting, Henry was considerably aided by his two assistants, Azizul Haque and Hem Chandra Bose to the extent that Haque was the originator of the classificatory and searching system although Henry did not acknowledge this at the time (Adam 2016, 109). Henry had originally sought out a local expert in mathematics and statistics from the local college and had recruited Haque to help make the anthropometric measurement system work (Beavan 2002, 137). However, Bertillon's system was notoriously difficult to use and met with only limited success. This prompted Haque to consider fingerprinting and to experiment with ten-digit data sets. It was apparently Haque's enthusiasm for fingerprinting which prompted Henry to begin corresponding with Galton. In England on leave in 1894, Henry visited Galton and agreed to collaborate. Galton was to keep Henry up to date with the fingerprint classification system which he was developing and Henry was to supply Galton with the fingerprints of convicts (Beavan 2002, 137–138).

However, Haque was very disappointed with Galton's (1895) book on a fingerprint classification system. Galton's system was highly complicated, involving counting ridge patterns at near microscopic level and distributing them over twenty-eight separate (subjective) categories. Galton seemed to have little understanding of the kind of operator who would be using a fingerprint system—clerks with a relatively modest education rather than highly trained scientists (Beavan 2002, 138–139). This was a piece of western technological knowledge that would not translate to an eastern setting. Haque developed his own ingenious and much easier to use system which involved subclassification of prints into sub-categories—physically arranged in a database of pigeonholes (see Beavan 2002, 139–141 for details). The system was far less error-prone than Galton's version and much easier to use than the older anthropometric system. Henry immediately saw its potential to advance his career. He put it about that the system had come to him in a flash of inspiration on a train journey, rather than admitting that it had been devised by Haque (Beavan 2002, 146).

Meanwhile, in 1897, India became the first country of the British Empire to adopt fingerprinting for criminal identification (Beavan 2002, 142). Henry's failure to acknowledge the roles of Haque and Bose in the early 1900s was quite shocking (Sodhi and Kaur 2005). As Sodhi and Kaur point out, there was little Haque and Bose could do about it at the time: "In 1900, native Sub-Inspectors of Police had no channel of redressal against a British Inspector General nor would a high official doubt that a junior, and that too an Indian, could file a representation against him" (Sodhi and Kaur 2005, 185). Although Henry made a belated acknowledgment of their role in the 1920s, and Haque is now recognized as the developer of the system, two individuals who made the essential contribution to a system used in criminal justice the world over (although later achieving some recognition for their efforts) were left in relative poverty in retirement and were never able to benefit appropriately from their innovations.

The information technology requirements of a fingerprint cataloguing and matching system require a one-to-many search. The database of cards where fingerprint images were stored needed to be organized in a way to classify the cards efficiently so an operator could quickly home in on a group of cards where a match was likely to lie, thus cutting down significantly on a brute force search. The "Henry" system was adopted by Scotland Yard when Henry was recalled to London to be Assistant Commissioner of the CID and to set up the Fingerprint Bureau in 1901. The branch was instantly successful, uncovering the pseudonyms of many hundreds of recidivists in its first year of operation alone (Beavan 2002, 13). Although he did not invent the classificatory system, Henry had successfully brought it to the UK. The Henry system was also adopted by the FBI in 1924. Such was the longevity of the Henry system, essentially a system designed to address the problems of colonial administration, that the first computer systems for fingerprint identification introduced in the 1960s were based on it—a remarkably long reach for a technology developed in a colonial regime (Wayman et al. 2005, 26–28). Fingerprinting thus provides an excellent example of a biometric identification technology that flowed from east to west, from a contact zone where local and western knowledge circulated.

6. Continental Criminalistics in Colonial Clothing

Other aspects of criminal identification technologies display a comparable pattern of colonial history and a similarly remarkable story of the movement of scientific and technological knowledge. My second example involves "criminalistics" or the scientific approach to the management and analysis of crime scenes and trace evidence. It includes procedures to link offenders to crimes and scenes of crime which was developed in the late nineteenth century in continental Europe and elsewhere. Broadly speaking, the term criminalistics incudes knowledge and techniques such as the correct way to preserve a crime scene, how to package and preserve material obtained from the crime scene, and the use of a wide range of scientific techniques such as microscopy and various chemical analyses to identify and analyze trace evidence.

Hans Gross is widely acknowledged as one of the major originators of criminalistics in the late nineteenth century. Having qualified as a lawyer in Graz, Gross then took up a position as an examining justice in Upper Styria (Austria) and later became a public prosecutor in Graz (Adam 2016, 65). Working in an inquisitorial legal system, his professional position as an examining magistrate enabled him to develop a wide knowledge of different crimes as not only was he responsible for taking evidence from witnesses, he was also responsible for managing the scenes of crime, including linking a potential offender to a crime scene through the examination of trace evidence. The publication of his *Handbuch für Untersuchungsrichter als Systeme der Kriminalistik* (*Handbook for Examining Magistrates as a System of Criminalistics*), first published in 1893, demonstrated a remarkably sophisticated understanding of a wide spectrum of crime, the means of identifying criminals and the problems of witnessing (Gross 1893; Adam 2016, 67). The *Handbuch* was written in German and it ran to six more German language editions with the final edition published in 1943. The first English language translation and *adaptation* was published in 1906 with four more English language editions published, the final edition in 1962 (Adam and Adam 1906).

Thus far, the story of the publication and translation into English of Gross's *Handbuch* appears unremarkable. However, when we consider the journey it made in relation to the publication of its English editions, we see that it undertook a lengthy journey from continental Europe to British India and then back to the UK. In fact, we can characterize the system outlined in the book which arrived in the UK in the early years of the twentieth century not so much as continental criminalistics, but as continental criminalistics in colonial clothing. Gross's original book was, of course, published in German. It was therefore inaccessible to most of the British criminal justice community and made little impact in the UK until an English language version was available. It is significant that the first translation was not undertaken in the UK, rather Gross's system arrived in the UK via a translation *and* adaptation undertaken in her most important colony.

John Adam, Barrister-at-Law and Crown and Public Prosecutor in the High Court at Madras and his son John Collyer Adam, Barrister-at-Law and Advocate also in the High Court at Madras, were the authors who produced the first translation and adaptation, published in 1906 in India and reprinted in London the following year (Adam and Adam 1906; Adam and Adam 1907). A second edition was published in London by John Collyer Adam in 1924 (Adam 1924). Subsequent editions were edited by senior UK police officers and published in the UK (Kendal 1934). There were considerable variations and amendments to the various editions reflecting changes in policing in the empire and in the UK. Criminalistic knowledge, originally developed in continental Europe, traveled to colonial India to be reworked and revised and was then transported to the UK, to be revised and remade yet again.

As translations and adaptations of Gross's work, the English editions were never *simply* translations, although they always badged themselves as such, possibly for reasons of prestige. It difficult to know how much of Gross's writing was actually left in them, considering that several new German editions appeared after 1903, running along a parallel publication track to the English language versions, and given that the English editions varied considerably from one to another. The English editions are best regarded as a palimpsest where the 1904 edition of Gross's work was translated and adapted in 1906, and then reworked and rewritten to suit different and evolving criminal justice settings.

Adam and Adam produced their edition translated and adapted for the Indian and colonial officials and police officers in the criminal justice system where they practiced (Adam and Adam 1906). Colonial policing was highly centralized in the Madras Presidency and the magistrate-collector was an administrative role responsible for collecting revenue, supervising the police and running the district courts (Arnold 1985). Hence there was a strong administrative aspect to their work, running hand in hand with legal duties, which parallels the use of fingerprinting for administrative purposes in India. When Adam and Adam claimed to have brought Gross's 1904 edition up to date in their 1906 work, they were referring to its reworking for colonial criminal justice purposes, rather than claiming that Gross's work had become out of date, in a technical or scientific sense, in a short space of time. Many of the examples and illustrations reflected this colonial reworking. The Madras government contributed various drawings of material from the chief constable, a catalogue of weapons was supplied by the Indian Museum in London (UK), illustrations were provided by a Madras gunsmith and a Madras publisher, botanical drawings were made from specimens supplied by Lt Col Van Geyzel, chemical examiner to the government of Madras (Adam and Adam 1907, vii). The book was, therefore, an important focus for a contact zone centering on and creating criminalistic knowledge, where such knowledge circulated from Continental Europe to colonial India and back to the UK.

There were further reasons why the first two English editions of Gross's work were edited for the criminal justice system in colonial India, beyond simply the provision of local examples. Two of the major themes of Gross's *Handbuch* were the demonizing of wandering tribes or gypsies and the question of unreliability of witnesses, even expert witnesses. Gross held very negative views of gypsies, believing them to be a criminal underclass forming a separate evolutionary branch to that of respectable Europeans (Adam 2016, 69). We have already noted that part of the spur for the introduction of fingerprinting in colonial India centered on suspicion of indigenous people, and the view that they were not to be trusted and that impersonation was rife. The question of trust raised itself in the colonial translation of Gross's work, as it was relatively easy to project his views on gypsies onto nomadic Indian tribes who were already treated with considerable suspicion by the British authorities. Adam and Adam regarded the traditional "gipsy industries" of mat and basket making as a cover for "nefarious practices" and regarded nomadic people as responsible for much of the crime in the country (Adam and Adam 1907, 355). Although there were only a few specific references to nomadic groups in Adam and Adam's work, nevertheless extensive descriptions of the habits, superstitions and activities of criminal groups were clearly meant to be projected on, and thus to demonize nomadic people. The British authority's approach involved criminalizing and at the same time trying to reform wandering tribes who were regarded as a threat, as they were unwilling to follow settled "respectable" occupations.

In the Madras Presidency where Adam and Adam worked, the Yerukulas were a significant wandering tribe who were officially declared a criminal tribe in 1913 under a revised version (1911) of the 1871 Criminal Tribes Act (Radhakrishna 2001, 27). The Yerukulas were successfully "reformed" by the Salvation Army into a settled existence providing cheap labor to work in a tobacco factory, thus serving British economic policies. Hence any threat that the Yerukulas posed was seen just as much in terms of commercial interests as concerns over crime. Indeed, the process of criminalizing and resettling the Yerukulas was so successful that apparently many contemporary Yerukulas regard their ancestors as criminals (Radhakrishna 2001). The British managed to superimpose their traditional distrust of gypsies onto the nomadic peoples of India, and

Adam and Adam's book can be seen as another part of official attempts to criminalize indigenous people. They were able to rework an important book from Continental Europe on criminalistics and the technologies of criminal identification, with its demonizing of "gypsies," to import criminalistic ideas on how "criminal tribes" should be treated to the information order of early twentieth century colonial India.

There is a further, final way in which trust or lack of it is represented in Adam and Adam's book. This relates more directly to the question of gentlemanly trust of scientific and technological witnesses, an important aspect of the movement of knowledge. Working in senior roles in the criminal justice system, Adam and Adam were naturally concerned with the reliability of witnesses, although this was surely a problem everywhere. Nevertheless, the colonial setting appeared to present particular problems of witness reliability in relation to expert witnesses. In India experts "of the first rank" were very seldom available. According to the Adams, the kind of experts who knew how to give evidence properly included: "Chemical Examiners to Government, the Lecturers on Medical Jurisprudence at Headquarters, the Government Experts in Finger-prints" (Adam and Adam 1907, xxii). They were particularly scathing about lower grades of hospital doctors and apothecaries, given that hospital assistants were often asked to conduct post-mortems and were far more likely to jump to rash conclusions in a criminal case than a more experienced medical official.

The Adams' concerns over the abilities or otherwise of expert witnesses may not look very different from similar problems in other parts of the world, but when they asserted that no "non-gazetted medical officer"[3] should be permitted to undertake a post-mortem in a suspicious case and report on it, they were making a very clear distinction between experts who could be trusted—namely senior officials in laboratories and government departments—and lower-graded medical doctors and local people in the strict hierarchy which existed in Indian society. They were able to make such a clear distinction on where trust should be placed because of the highly structured Indian system of gazetted and non-gazetted officials. This allowed them to make an easy distinction between officials in senior positions who were mainly European and who could be regarded as trustworthy, and those who were in junior positions who were mainly indigenous people who were not to be regarded as trustworthy. For instance, a list of Official Chemical Appointments for the UK and her colonies in 1906 confirms that, in India at that time, senior appointments were held by people with European, mainly British, names. Indian names, when they appear in this list were usually in "Assistant" roles (Pilcher 1906).

The third edition of the English edition of Gross's *Handbook* was edited by Norman Kendal, CBE, Assistant Commissioner of the Criminal Investigation Department, Metropolitan Police in 1934 (Kendal 1934). In this edition, Indian examples had been replaced by examples from leading forensic experts from the UK. Kendal firmly positioned the book as a practical textbook for the burgeoning interest in scientific detection in the UK, and the development of forensic science laboratories in England from the 1930s. There was much less emphasis on demonizing "gypsy" groups in this edition, which was arguably not a significant issue in the UK in the 1930s. Unsurprisingly, references to "gypsy" groups were dropped completely in later editions published after World War II.

In this way, Hans Gross's remarkable treatise on the new science of criminalistics made its journey from continental Europe, to British India and then to the UK whilst being translated, rewritten and remade to include indigenous knowledge, to reinforce ways in which criminals could be identified and linked to crimes and crime scenes, and to emphasize hierarchies of who was to be trusted in the making of scientific, technological and legal knowledge. The knowledge traveled not because it was "true" as in the old modernist versions of the diffusion of scientific and technical knowledge. Rather it traveled because it could be rewritten and revised to suit the local needs of administrative and criminal justice systems in different parts of the world.

7. Conclusion

The fingerprinting and criminalistics handbook examples display a number of clear congruences with Raj's postcolonial analysis of the movement and circulation of scientific and technical knowledge. Both examples show how knowledge mutated to suit the purposes of the administrative regimes where technical knowledge was developed for particular civil and criminal justice purposes, based on economic interests and was taken up and developed further in maintaining social order and blending British colonial codes and procedures with local processes already in place. In these important contact zones, indigenous people were not passive informants, rather they were active in the making of knowledge, although the relationship was often unequal.

The geographical turn in the philosophy of science and technology drawing on postcolonial theory is now firmly in process. There are a number of excellent case studies to act as exemplars for philosophy of technology to develop its "postcolonial turn" further (e.g. see Law and Lin 2017; Raj 2007). As is so often the case, these analyses show that philosophy of science and technology is best made in relation to historical examples, nowhere more so than in relation to understanding the movement of scientific and technological knowledge, for which the "information order" of fin-de-siècle India provides significant examples.

A historical postcolonial analysis in the philosophy of technology has much to offer contemporary philosophy of technology in relation to current concerns over biometric identification and surveillance technologies, some of which can be regarded as twenty-first century descendants of earlier identification technologies such as fingerprinting. In the historical fingerprinting and criminalistics examples described previously, I noted that much of the push towards implementing these identification technologies rested on suspicion of indigenous, often nomadic, groups. Their way of life, the fact that they were nomadic rather than "settled," fueled colonial suspicions of criminality. The Yerukulas people were only regarded as successfully "reformed" by the Salvation Army when they were settled in camps and employed to work on its land, thus providing cheap regular labor. This suited British economic policies which were designed to raise revenue, and against which wandering people had traditionally been regarded as a threat (Adam 2016, 69). In India the interests of British overseas trade were served by the enormous edifice of the East India Company (Raj 2007, 107). Indeed after the conquest of Bengal in the late eighteenth century subsequent administrative governance, including the administration of civil and criminal law, was firmly in the hands of the East India Company (Raj 2007, 108–109). The Yerukulas example serves to reinforce the point that state power and corporate power are thoroughly intertwined, arguably a historical illustration of "surveillance capitalism" (Zuboff 2019).

Although the following, contemporary example is not set in a colonial space in quite the same way, nevertheless the oppression of a minority group by the machinery of state through the implementation of identification technologies, has strong parallels with the historical examples described earlier in this chapter. Taking face recognition technologies as a contemporary example, it is clear that these technologies are being used to negative effect in parts of the world.

Anthropologist Darren Byler's (2019) study of the use of face recognition technology by the Chinese authorities to "pioneer a new form of terror capitalism" against the Uyghur people, an Islamic faith group, makes for very chilling reading. As Byler (2019) describes, in its "People's War on Terror," the Chinese government treats most expressions of Uyghur Islamic faith as religious extremism, ethnic separatism and even potential terrorism, for which detention centers have been set up for supposed offenders. Religious and political activity is discovered through checks on social media activity on Uyghurs' smart phones at the many checkpoints in the Xinjiang region of north-west China and these checks are used to put Uyghurs in detention centers (Byler 2019).

The Integrated Joint Operations Platform is a regional data system that uses AI to monitor checkpoints around Xinjang's cities so that attempts to enter public buildings and spaces such as banks, hospitals, shopping centers or parks will be monitored (Byler 2019). This means that biometric and data histories are being used to keep people in place, often literally, at home. Face recognition software is a key part of this technological armory. Face scanners check the Uyghurs' movements; automated face recognition is a significant part of the story of using technology to monitor and control. Videos are scanned to see who is acting suspiciously and who may become "unsafe"—this involves activities such as dressing in an Islamic manner. The war on terror offers commercial companies in the service of the state opportunities to build and experiment with systems which can monitor and control people's lives, an extreme version of Zuboff's (2019) "surveillance capitalism." As Uyghurs began to successfully use new mobile technologies a decade or more ago to promote their expressions of faith and identity, and even to protest against the violation of their human rights, so too did the state begin to use new technologies to clamp down on such expressions. Surveillance technologies are used in attempts to predict criminal behavior and to repress groups of people, including Uyghurs who are pushed into "retraining" to be politically docile yet economically productive in ways that the state prescribes (Byler 2019). This resonates with the attempts of the British to "retrain" the Yerukulas people in colonial India to become settled and economically productive.

One of the main software tools used by the Chinese state in its oppressive surveillance of Uyghurs is an AI system which can automate identification of Uyghur faces based on shape and color of facial features. Biometric records, "face signatures," have been created by scanning individuals from different directions. Although, as Byler (2019) points out, there is considerable western skepticism about whether the AI face recognition software used in China can actually work, nevertheless thousands of Uyghurs have been detained as the result of this surveillance technology; the effects are very real. Unfortunately, this reinforces the lengthy history of ways in which colonial, state, and corporate use of biometric identification technologies have historically been, and continue to be, used for repressive ends.

Notes

1. See Anderson and Adams 2008, 182 for the "Marie-Celeste" model.
2. See Raj 2007, Chapter 3, 95–138.
3. Public officials appeared in the Gazette of India and were classified as "gazetted" if their rank was managerial or executive. Non-gazetted officials held middle or lower-ranking positions.

References

Adam, Alison. *A History of Forensic Science: British Beginnings in the Twentieth Century*. Abingdon and New York: Routledge, 2016.

Adam, John, and John Collyer Adam. *Criminal Investigation: A Practical Handbook for Magistrates, Police Officers and Lawyers*. Madras: A. Krishnamachari, 1906.

Adam, John, and John Collyer Adam. *Criminal Investigation: A Practical Handbook for Magistrates, Police Officers and Lawyers*. London: The Specialist Press, 1907 (reprint of 1906 edition).

Adam, John Collyer. *Criminal Investigation: A Practical Textbook for Magistrates, Police Officers and Lawyers*. Translated and adapted from the *System der Kriminalistik* of Dr. Hans Gross. London: Sweet and Maxwell, 1924.

Anderson, Warwick. "Introduction: Postcolonial Technoscience." *Social Studies of Science* 32, no. 5/6 (October–December, 2002): 643–658.

Anderson, Warwick, and Vincanne Adams. "Pramoyeda's Chickens: Postcolonial Studies of Technoscience." In *The Handbook of Science and Technology Studies*, edited by Edward J. Hackett, Olga Amsterdamska, Michael Lynch, and Judy Wajcman, 181–204. 3rd edition. Cambridge MA and London: MIT Press, 2008.

Arnold, David. "Crime and Crime Control in Madras, 1858–1947." In *Crime and Criminality in British India*, edited by Anand A. Yang, 62–88. Tucson, University of Arizona Press, 1985.

Basalla, George. "The Spread of Western Science." *Science* 156 (May 1967): 611–622.

Beavan, Colin. *Fingerprints: Murder and the Race to Uncover the Science of Identity*. London: Fourth Estate, 2002.

Byler, Darren. “Ghost World.” *Logic: A Magazine about Technology*, 7 (May 1, 2019). https://logicmag.io/07-ghost-world/.

Cole, Simon A. *Suspect Identities: A History of Fingerprinting and Criminal Identification*. Cambridge, MA: Harvard University Press, 2001.

Foucault, Michel. *Power/Knowledge: Selected Interviews and Other Writings 1972–1977*. Brighton: Harvester, 1980.

Galton, Francis. *Finger Prints*. London and New York: Macmillan, 1892.

Galton, Francis. *Fingerprint Directories*. London and New York: Macmillan, 1895.

Gross, Hans. *Handbuch für Untersuchungsrichter als System der Kriminalistik*. 2 volumes. Munich: J. Schweitzer Verlag, 1893.

Harding, Sandra. *Is Science Multicultural? Postcolonialism, Feminisms and Epistemologies*. Bloomington and Indianapolis: Indiana University Press, 1998.

Harding, Sandra. *Sciences from Below: Feminisms, Postcolonialities, and Modernities*. Durham, NC and London: Duke University Press, 2008.

Harding, Sandra, ed. *The Postcolonial Science and Technology Studies Reader*. Durham, NC and London: Duke University Press, 2011.

Kendal, Norman. *Criminal Investigation: A Practical Textbook for Magistrates, Police Officers and Lawyers*. 3rd edition. London: Sweet and Maxwell, 1934.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

Latour, Bruno. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press, 1987.

Law, John, and John Hassard. *Actor Network Theory and After*. Oxford and Malden, MA: Blackwell Publishing, 1999.

Law, John, and Wen-Yuan Lin. “Provincializing STS: Postcoloniality, Symmetry, and Method.” *East Asian Science, Technology and Society* 11, no. 2 (2017): 211–227.

MacLeod, Roy. “On Visiting the ‘Moving Metropolis’: Reflections on the Architecture of Imperial Science.” *Historical Records of Australian Science* 5, no. 3 (1982): 1–16.

McNeil, Maureen. “Introduction: Postcolonial Technoscience.” *Science as Culture* 14, no. 2 (2005): 105–112.

Palladino, Paolo, and Michael Worboys. “Science and Imperialism.” *Isis* 84 (1993): 91–102.

Pilcher, Richard B. *A List of Official Chemical Appointments in Great Britain and Ireland, in India and the Colonies*. London: Institute of Chemistry, 1906.

Pratt, Mary Louise. *Imperial Eyes: Travel Writing and Transculturation*. London and New York: Routledge, 1992.

Radhakrishna, Meena. *Dishonoured by History: “Criminal Tribes” and British Colonial Policy*. Hyderabad: Orient Longman, 2001.

Raj, Kapil. *Relocating Modern Science: Circulation and the Construction of Knowledge in South Asia and Europe, 1650–1900*. Houndmills, Basingstoke: Palgrave Macmillan, 2007.

Said, Edward. *Orientalism*. New York, NY: Pantheon, 1978.

Schiebinger, Londa. “Forum Introduction: The European Colonial Science Complex.” *Isis* 96 (2005): 52–55.

Sengoopta, Chandak. *Imprint of the Raj: How Fingerprinting Was Born in Colonial India*. Basingstoke and Oxford: Pan, 2003.

Shapin, Steven. *A Social History of Truth: Civility and Science in Seventeenth Century England*. Chicago, IL: Chicago University Press, 1994.

Shapin, Steven. 1998. "Placing the View from Nowhere: Historical and Sociological Problems in the Location of Science." *Transactions of the Institute of British Geographers* 23, no. 6 (1998): 5–12.

Sodhi, G. S., and Kaur, Jasjeet. "The Forgotten Indian Pioneers of Fingerprint Science." *Current Science* 88, no. 1 (January 2005): 185–191.

Spivak, Gayatri Chakravorty. "Can the Subaltern Speak." In *Colonial Discourse and Post-Colonial Theory: A Reader*, edited by Patrick Williams and Laura Chrisman, 66–111. New York: Columbia University Press, 1993.

Terrall, Mary. "Heroic Narratives of Quest and Discovery." In *The Postcolonial Science and Technology Studies Reader*, edited by Sandra Harding 84–102. Durham, NC: Duke University Press, 2011.

Thomas, Nicholas. *Colonialism's Culture: Anthropology, Travel and Government*. Princeton, NJ: Princeton University Press, 1994.

Traweek, Sharon. *Beam Times and Life Times*. Cambridge, MA: MIT Press, 1988.

Wayman, James, Anil Jain, Davide Maltoni, and Dario Maio, eds. 2005. *Biometric Systems: Technology, Design and Performance Evaluation*. London: Springer-Verlag.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile, 2019.

CHAPTER 12

RAWLS, INFORMATION TECHNOLOGY, AND THE SOCIOTECHNICAL BASES OF SELF-RESPECT

ANNA LAUREN HOFFMANN

1. Introduction

The work of political philosopher John Rawls has featured prominently in discussions of information, technology, and ethics (see Hoffmann 2017). However, the vast majority of these efforts overlook the substantive and justificatory role of what Rawls (1971) calls the social bases of self-respect, which he counts as "perhaps the most important" (386) of the primary goods his two principles of justice are designed to distribute. In some ways, this lack of work on self-respect is reflective of a broader absence of consideration paid to respect in information and computing ethics, as lamented by Dillon (2010). But the development and exercise of self-respect is, like other important human values, shaped by the affordances and moral valences of technology in ways that merit particular and sustained attention.

In the following sections, I attend to the role of self-respect as it relates to issues of social justice, information, and technology. Beginning with Rawls' work, I detail the importance of self-respect for theories of justice generally while also moving past his individualist conception in favor of a social understanding of self-respect informed by race-based, feminist, and leftist work. This expanded notion of self-respect emphasizes its social contingency—that is, the ways self-respect is not only a matter of individual motivation, but also fundamentally shaped by social, political, and economic conditions. After establishing the importance of self-respect, I draw on work in both values-conscious design and disability studies to show how self-respect can also be promoted or undermined by the design, dissemination, and use of technology. More precisely, I argue that the sociotechnical relationships supported by, in particular, information technology play an important role in codifying, entrenching, and reproducing self-respect's social bases. From there, I deploy Wolff's (1998) notion of "respect-standing" as a heuristic for uncovering information technology's impact on self-respect in two domains: (1) privacy and surveillance and (2) information and identity. In doing so, I demonstrate how a move from *the social bases of self-respect* to *the sociotechnical bases of self-respect* can help us better account for self-respect in ethical analyses of technology.

2. Rawls and the Social Bases of Self-Respect

According to Rawls, the social bases of self-respect are integral to the development of what he calls *the two moral powers*, defined as capacities to (1) recognize and act from justice's demands and (2) adopt and take up effective means to some more or less complete set of valued ends. In view of this, Rawls lists the "social bases of self-respect" as among the primary goods his theory of justice is designed to distribute, even going so far as to call it "perhaps the most important primary good" (Rawls 1971, 386). As a primary good, the social bases of self-respect provide an individual with both "a sense of his own value" and a "secure conviction that his conception of his good, his plan of life is worth carrying out" (Rawls 1971, 386). Rawls' use of the masculine pronoun aside, we see that the first aspect of self-respect affirms the value of individuals' plans of life, while the second affords individuals a confidence necessary to those plans (Zink 2011, 332). In this way, the social bases of self-respect are integral to the effective exercise of the capacity to set and pursue a conception of the good—that is, Rawls' second moral power.

Elsewhere, Rawls connects self-respect to the first moral power during his *argument from stability*. Rawls believes that not only should a conception of justice be justifiable to parties in the original position, but it should also be stable—that is, it ought to cultivate in individuals a sense of justice and discourage countervailing inclinations or attitudes (Zink 2011, 338). In particular, a conception of justice should promote values like self-respect and discourage tendencies towards envy or resentment that, over time, might undermine the development of Rawls' first moral power. For parties selecting principles of justice in the original position, if one conception of justice better promotes this moral power (by, among other things, supporting the development of self-respect) then it is said to be more stable—and stability counts as a reason for parties to choose that conception.

Rawls argues that the lexical ordering of his two principles of justice—that is, his requirement that the first principle (the liberty principle) be satisfied prior to the second (the opportunity principle)—offers more stability than principles from other philosophical traditions. First, Rawls believes that his prioritization of liberty helps individuals cultivate an effective sense of justice (i.e., Rawls' first moral power) and better supports their self-respect. As Cohen (2003) summarizes, self-respect is, on Rawls' account, most stable when rooted on one's sense of oneself as an equal member of society, sharing responsibility for making fundamental judgements about social and political issues (109). Second, he argues that second principle considerations (fair equality of opportunity and the difference principle) support individuals' relative socioeconomic independence, ensuring no one must be wholly subservient to another—a condition that would be detrimental to one's self-respect.

Combined, these two features of Rawls' work—the social bases of self-respect and the argument from stability—show how self-respect is integral to his theory of justice. It also demonstrates the foundational role self-respect plays in establishing and stabilizing egalitarian social arrangements, since it supports individuals' sense of equal membership in society (Mathiesen 2015, 440). In this way, his work establishes the importance of self-respect for the stability of liberal egalitarian theories of justice generally. At the same time, it also exposes some limits of Rawls' conception of self-respect. Rawls clearly views self-respect as "a matter of individual motivation" and that those who lack it "do not possess the psychological disposition necessary for acting from a sense of justice" (Zink 2011, 338–339; see, also: Rawls 1971, 440–446; Dillon 1997, footnote 18, 232). But Rawls' two principles of justice do not exhaust the social and cultural sources that may be relevant to the development of self-respect in individuals, especially forms of stigmatization, disdain, or humiliation (Young 1990; Young 2006; Pilapil 2014).

This is not to say that social considerations are wholly absent from Rawls' account. He notes, for example, that maintaining a sense of one's value "depends in part upon the respect shown to us by others; no one can long possess an assurance of his own value in the face of enduring contempt or even the indifference of others" (Rawls 1999, 171). Here, self-respect, while still fundamentally rooted in the individual, is contingent on the recognition that one is seen as a fully cooperating member of society (Rawls 1993, 318). Further, Rawls argues in his characterization of the family—in line with liberal theory generally—that the home is a uniquely intimate sphere of personal development and that a theory of justice must not unduly intrude on its inner-workings. Given the relationship between self-respect and his second moral power, it's clear that the family plays an important role in the development of individuals' self-respect.

Despite these gestures, his individualist conception of self-respect generates some lingering problems. Eyal (2005), for example, argues that Rawls' characterization of self-respect ultimately commits him to objectionable or even illiberal politics, as his commitment to individualistic self-respect as "perhaps the most important primary good" should logically force him to abandon the priority of liberty in favor of strict equality in self-respect's social bases. For others, Rawls' conception is less logically fatal; instead, it simply necessitates further explication of what might make up self-respect's "social bases" and whether or not those things are distributable in ways similar to, for example, income (Doppelt 2009, 128). His image of the (patriarchical) nuclear family, for example, abstracts away from the often oppressive realities of many family situations—realities that require attention to both unfair distributions of resources and misogynistic cultural norms.

For present purposes, however, I accept and affirm Rawls' insight that self-respect is not only important, but integral to the realization of social justice. Without a secure conviction in one's self and one's plan of life, moving through the world and pursuing one's valued ends is comparatively more difficult. However, accepting self-respect's value does not simultaneously mean adopting Rawls' views uncritically or without exception. Rather, we must take care to further articulate and extend our understanding of self-respect's social bases in order to better understand how it may be supported or undermined. If our aim is to ultimately move from the largely ideal realm of Rawls' work to achieving social justice under non-ideal conditions, then we need to be explicit about the social conventions and contexts that shape the development of self-respect today.

3. Taking Self-Respect's Social Bases Seriously

Self-respect's social dimensions have generated explicit philosophical discussion since at least the mid-twentieth century, both prior to and in conversation with Rawls' work. Telfer (1968), for example, argues that self-respect hinges on an independence from others (117)—though she does not specify the degree of independence required. Darwall (1977) makes self-respect's social contingency more explicit, noting that its realization depends, in part, "on the appropriate conception of persons and on what behaviors are taken to express this conception or the lack of it" and may "vary with society, convention, and context" (48). Attention to social convention matters as individuals' lives are informed by a range of contexts, from networks of friends and family to workplaces, neighborhoods, and nation-states (Doppelt 2009, 132). Each of these contexts can have profound and pervasive impacts on the possibilities for self-respect available to individuals and groups.

If self-respect is, in many ways, social, our analyses must pay close attention to the contours of those social frameworks and contexts that underwrite its development. For Dillon (1997), self-respect is profoundly shaped by our "basal self-understandings" that inform our moral development long before we begin to exercise agency. These basal frameworks "are constructed in the complex, emotionally charged interplay of self, others, and institutions which begins before we are capable of conceptualizing self, worth, persons, institutions, and the relations among them, and it shapes and delimits . . . our agentic capacities" (Dillon 1997, 244). In this way, self-respect is—at its base—constructed through the complex interplay of social, cultural, and political forces.

Importantly, the basal self-understandings that support our self-respect are, for some, forged within social contexts of oppression (Dillon 1997, 245–246). This presents particular problems for conceptions of self-respect as solely a kind of independence or matter of individual motivation, especially in cases of internalized oppression (see: Charles 2010). In the United States and elsewhere, individuals' basal frameworks are shaped by histories of colonialism, genocide of native peoples, slavery, discrimination and disenfranchisement, and other institutionalized injustices. As Moody-Adams (1993) argues, for example, the development and maintenance of self-respect for Black individuals is often constrained by normative standards of race embedded in social, political, and economic structures. Specifically, white hegemonic norms and expectations of appearance, behavior, and beyond create both explicit and implicit barriers for the development of self-respect. As poet and writer Morgan Parker (2017) captures it in her essay "How to Stay Sane While Black," "every time I tell myself that I am worthless, how do I know whether it's me thinking it, or the white voices I've internalized?" (para. 12).

To be certain, the presence of barriers does not make the development and exercise of self-respect impossible. It does, however, shape the conditions and means by which self-respect is realized and maintained. As Thomas (1995) and Boxill (1976; 1992) argue, for example, political protest during the American civil rights movement of the 1960s was not exclusively about the winning of specific rights for African-Americans—it was also an effort to liberate self-respect for marginalized Black communities generally. Their accounts follow Rawls in admitting the profound influence of social institutions on the development of self-respect, but they are more explicit in attending to the role of protest for transforming unjust institutional structures and asserting self-respect.

In addition to race and ethnicity, Rawls' heavily criticized characterization of the family reveals how the development of self-respect is also contingent on sex and gender. As Nussbaum (2004) notes, "the family is one of the most non-voluntary and pervasively influential of social institutions, and one of the most notorious homes of sex hierarchy, denial of equal opportunity, and sex-based violence and humiliation" (115). Though Rawls recognizes the equal standing of all family members as citizens, he fails to offer an appropriate response to injustices within the family's structure itself. This is insufficient, as the equal provision of the social bases of self-respect must take seriously issues of sex-based subordination and oppression both in the home and more broadly, as the development of self-respect is intimately tied to one's place within a larger culture and whether or not that culture forces particular social roles upon certain categories of people (Okin 2004, 202). It must also pay attention to the ways embedded heterosexist standards of sexuality and cisgender norms of binary gender shape the development of self-respect for lesbian, gay, and bisexual individuals (Mohr 1988) and transgender, intersex, or gender non-conforming individuals, respectively.

Finally, self-respect is also often informed by conditions of work and employment—especially by uneven distributions of decision-making power that structure socioeconomic relations. As Doppelt (1981) argues, "Rawls' conception does not adequately comprehend . . . the deep ways in which equality and inequality in its social bases are decisively shaped by the distribution of economic power and position in advanced industrial society" (260). As Rawls (2007) himself points out in his lectures on Marx, leftist conceptions are suspicious of the assumption that the conditions under which individuals are able to exercise certain moral ideals can be improved independent of economic circumstances. On this account, the realization of self-respect for certain individuals (workers) is unduly subject to the decisions of others (capitalists) that drive economic relations. These individuals are constantly subject, as Marx (1975) put it, to the "whims of the wealthy" (283).

4. Sociotechnical Relations and Self-Respect

The preceding discussions lay bare the ways self-respect is more than solely a matter of individual motivation. But even the more expansive, social accounts of self-respect fail to describe how material artifacts and practices work to entrench social and political norms, persisting and shaping individuals' experiences over time. Put another way, an emphasis on the *social* overlooks the role of technology and *sociotechnical relations*—that is, relations defined by the "combinations of hardware and people (and usually other elements) to accomplish tasks that humans cannot perform unaided by such systems" (Kline 2003, 211)—in constituting and entrenching the social bases of self-respect in both material and practical ways. Importantly, our self-respect is not won or lost only in our interactions with others; it is also shaped by our interactions with non-human dimensions of the world—like technological artifacts, information systems, and the built environment—that codify and reproduce self-respect's social bases.

Choices made during the conception, development, and dissemination of technological artifacts and systems imbue them with particular values; at the same time, those built in values press on users and the world and, subsequently, further inform the shape of human values. Consequently, technology does not passively mediate, but actively shapes our moral, political, and cultural development (Verbeek 2009). Our moral analyses, then, should attend to the ways in which the design and development of technological artifacts and information systems might promote or obscure different moral values or ethical norms (Brey 2010, 41–42).Work in the area of values-conscious design (see, for example: Friedman and Nissenbaum 1996; Friedman, Kahn, and Borning 2006; Flanagan et al., 2008), in particular, is driven by a "concern over the moral and ethical consequences of our modern technological era" and focuses on ways to "ensure that particular attention to moral and technical values becomes an integral part of the conception, design, and development" of technology (Manders-Huits and Zimmer 2009, 38).

The moral valences built into technology (Verbeek 2009) can, along with the broader social structures within which they are deployed, have a profound impact on individual possibilities for the development and exercise of self-respect. As Brey (2007) describes, "the same technological artifact may empower one user more than it does another [since] artifacts will necessarily serve certain goals or interests better than others [and] may be more or less compatible with the attributes of users" (17). And although any single artifact or system cannot account for every possible user, there is—as Wittkower (2016) points out—a point where exclusion crosses over from pragmatically necessary to discriminatory, especially when interpreted in the appropriate social and historical context. Patterns of disempowerment, exclusion, and discrimination built (knowingly or incidentally) into technological artifacts and systems work to systematically hinder the development of self-respect for some, while promoting (or at least not standing in the way of) its realization for others.

The relationship between technology design and self-respect is made explicit in discussions surrounding disability. As disabilities activism and scholarship has shown, what counts as a disability is often determined not by any particular abilities exhibited by persons but, rather, by features of the social and physical environment (Oliver 1981; Shakespeare 2010; Barnes 2012). In this way, disability is something that is "imposed on top of" physical or other impairments (UPIAS 1976). For example, blindness is only a disability with regard to reading in the absence of Braille; similarly, being wheelchair-bound is only a disability with regard to mobility in the absence of accessible buildings. Further, as Shew (2017) points out, disabilities hinge not only on the presence or absence of assistive or accommodating technologies, but also on their maintenance and the social meanings attached to them (n.p.; see also Bell 2010; Docherty, et al. 2010). With regard to the latter, Terzi (2010) notes that persons with disabilities face difficulties "in dealing with the reactions by other people to the way they look, act, or simply to the way they are" (163), the complexities of which have been explored by Garland-Thomson (2006). Social attitudes and circumstances, then, "question disabled people's equal social bases of self-respect" (Terzi 2010, 163).

Building on these insights, the remainder of this section explores the sociotechnical bases of self-respect in two overlapping, but distinct areas of concern: (1) privacy and surveillance and (2) information and identity. I show how the affordances, norms, and assumptions "baked in" to the design, dissemination, and use of, in particular, information technology work to create differential conditions for the development of self-respect for different groups of people. To be clear, the point is not to show that such conditions will always, without regard to other factors, contribute to the diminishment of self-respect. Rather, I only mean to show how it might be that sociotechnical factors are complicit in the promotion of the self-respect of some while undermining it for others.

5. Analyzing the Sociotechnical Bases of Self-Respect

In order to see how self-respect's social bases are produced, reproduced, and codified through information technology, it will be helpful to first have some sort of heuristic or guide to identifying some of the ways technology might invoke self-respect. To this end, Jonathan Wolff's (1998) notion of "respect-standing" presents one concrete way to think about the ways social, political, or other forces may work to undermine self-respect. On Wolff's (1998) account, a person's respect-standing is defined as the degree of respect others have for that person (107). If individuals are treated with contempt, they will likely be led to believe that they have low respect-standing; conversely, if individuals are treated decently, they will likely believe their respect-standing is high (Wolff 1998, 107). When paired with Dillon's (1997) argument that pervasive subordination or devaluation of a category of persons can impact the respect persons can have for themselves, the notion of "respect-standing" helps us identify patterns of contempt (or, conversely, decent treatment) that inform the development of self-respect.

Wolff describes three ways in which one's respect-standing might be (reasonably or unreasonably) diminished: failures of common courtesy, mistrust, and shameful revelation. Failures of courtesy address situations where one is frequently ignored, patronized, or lectured, leading one to believe that she has low-respect standing (Wolff 1998, 108). In the workplace, for example, women have described situations wherein their ideas or contributions are not "heard" by others until they are repeated or reiterated by a colleague who is a man, often without attribution (Dodgson 2018). This phenomenon—colloquially known as "hepeating," a play on "repeating" (Gugliucci 2017)—is indicative of an uneven social distribution of respect. Where one category of persons (in this case, women) must struggle to be heard in ways other categories of persons (in this case, men) do not, we can expect the development and maintenance of self-respect and a sense of one's worth to be more emotionally or psychologically laborious for the former than for the latter.

Similarly, systematic patterns of mistrust can also undermine the respect-standing of entire categories of persons. Being asked to justify oneself or being called to account too often, or when similarly situated others are not, or when the depth of investigation seems out of proportion, is insulting—it gives the impression that one is not trusted, that one is an object of suspicion and is not being respected (Wolff 1998, 108). Here, persons' respect-standing can be undermined by uneven patterns of trust in society—as when some are subject to disproportionate and invasive investigations or are made to account for their day-to-day actions or beliefs more often than others. "Broken windows" policing policies, for example, intentionally skew law enforcement resources toward so-called "quality of life" offenses like vandalism or public drinking. Of course, the ideal "quality of life" often encodes particular racial or class biases, often privileging affluent and largely white standards of decorum or appearance. So, while the practice superficially appears not to target specific groups of people, like those of low socioeconomic standing or of minority racial or ethnic groups, its effect in practice is to subject these groups to increased surveillance and outsized levels of policing.

Finally, Wolff's (1998) third source of diminished respect-standing involves what he calls "shameful revelation" (109–110). In instances of shameful revelation, one is forced to behave in a certain way or reveal things about themselves that reduce their respect-standing (Wolff 1998, 109). Specifically, people are forced to reveal details about themselves or their lives that may be perceived as embarrassing or shameful. Even if there is no good reason why a particular trait should lower your respect-standing, it can still be experienced as a source of shame (Wolff 1998, 114–115). Consider, for example, the practice of "outing" lesbian, gay, bisexual, transgender, or queer individuals. Though activism and other efforts have, in the US context, made some progress towards lessening the shame and stigma attached to LGBTQ, acceptance and safety are far from evenly or consistently distributed. The practice of "outing" retains its force, in part, because normative background assumptions about sexuality or binary gender identity still work to structure LGBTQ identities as, at best, "other" or different and, at worst, deviant or shameful. This is particularly true for transgender, intersex, and gender-nonconforming individuals, who continue to face violence and harassment at greater rates than, for example, white and affluent cisgender gays and lesbians.

In the following two domains, I trace these three mechanisms—failures of courtesy, systematic mistrust, and shameful revelation—and their manifestation by and through the sociotechnical bases of self-respect. In each domain, the design and affordances of information technology conspire with existing patterns of social contempt and injustice to produce differential treatment for different groups of people. In doing so, they demonstrate how a move from *the social bases of self-respect* to *the sociotechnical bases of self-respect* can help us better account for the relationship between self-respect and technology.

5.1 Domain 1: Privacy and Surveillance

The values of respect and privacy have long been bound up with advances in information technology. Warren and Brandeis's (Warren and Brandeis 1890) paradigmatic framing of privacy as "the right to be let alone," for example, was a direct response to the increased popularity of Eastman Kodak Company's small and inexpensive snap cameras, which allowed almost anyone to become a photographer and further propagated salacious gossip papers (Solove 2010, 15). While Warren and Brandeis did not use the language of self-respect specifically, they nonetheless sought to affirm the fundamental role of privacy in preventing indignities and securing "the protection of the person" Subsequent claims to privacy made against technological invasions have followed this logic, also appealing to ideals of individual autonomy, self-determination, and dignity (Westin 1967; Benn 1971; Schoeman 1984). Reminiscent of Rawls' defense of self-respect, Regan (1995) argues that "privacy inheres in the individual as an individual and is important to the individual primarily for self-development or for the establishment of intimate or human relationships" (24). Similarly, Bloustein (1984) describes privacy as preserving an "individual's independence, dignity, and integrity; it defines man's essence as a unique and self-determining being" (163).

On these accounts, privacy is one means by which we respect individual dignity and, by extension, provide an individual with a sense of their own value constitutive of self-respect. In particular, privacy helps to cordon off and preserve spaces where, as Julie Cohen (2012) notes, individuals are free to "play"—socially, morally, culturally—and explore our identities, values, goals, and, ideals. Here, privacy is one means by which we can connect Rawls' second moral power and self-respect, since private reflection and exploration of different identities and plans of life is integral to developing a conception of the good upon which self-respect rests. Further, as Shannon Vallor (2016) argues, surveillance technologies that eliminate or degrade these private spaces "may shortchange our moral and cultural growth in the long term" (191).

But privacy, as with self-respect, cannot be fully accounted for by discussions of the individual. As Reiman (1976) points out, privacy is integral to establishing and maintaining intimate human relationships. In a different way, Nissenbaum (2010) connects privacy and the social through the notion of *contextual integrity*. She argues that social context is characterized, in part, by "context-relative informational norms," as she describes them, that "prescribe, for a given context, the types of information, the parties who are the subjects of the information as well as those who are sending and receiving it, and the principles under which this information is transmitted" (Nissenbaum 2010, 141). Privacy violations occur when the norms that govern the flow of personal information in a given context are upset in certain ways.

These "context-relative informational norms" have long been shaped by the affordances of available technologies of information production, storage, and dissemination. As Braman (2006) describes, many contexts—especially liberal bureaucratic ones—require the collection and processing of vast amounts of information in order to function (33–34). This collection and processing of information in the abstract hinges not only on the social expectations articulated by Nissenbaum, but also on the availability and use of material artifacts (paper, file cabinets, hard drives, networked computers) and the deployment of particular schematic practices (classification systems, organizational schemes). These artifacts and practices are not merely instrumental, but constitutive of one's understanding of given informational norms. For example, my expectation that sensitive information about me recorded on paper and shared with a third party will be kept confidential is determined not only by my trust in the third party, but also by the presence (or absence) of the material means for security, like a locked file cabinet. In this way, information technology is an integral part of the social bases of self-respect.

Today, online platforms like social networking sites take up much of the work of developing and regulating norms of information exchange. Despite the "open, neutral, egalitarian and progressive" connotation of the term "platform," however, these services are not neutral conduits for information exchange (Gillespie 2010). They are, instead, engaged in various forms of social, political, and economic mediation of online content (Klonick 2017; Gillespie 2018; Roberts 2018). Using a combination of human labor and computer software, online platforms actively set and inform the conditions and rules under which information can be shared, even if such interventions are, at times, hard to see (Gillespie 2010, p. 358). This kind of pervasive informational (and often algorithmic) gatekeeping raises important questions around fairness and transparency (Suzor 2018), democratic participation (Vaidyanathan 2018), and the role of computational agency in social and economic life (Tufekci 2015).

Platforms' design choices can have a profound impact on the informational norms and privacy expectations of users. For example, the introduction of Facebook's NewsFeed in 2006—an algorithmically curated stream of updates and advertisements based on a users' network of friends, interests, and engagement—shifted the flow of information within the service from the manual navigation of static profile pages to an automated stream of user updates visible upon logging into the site. This shift "threatened the privacy of users who previously assumed that only those friends who happened to visit their page would notice the changes; instead, any change made was automatically fed to all followers" (Zimmer and Hoffmann 2011, 177). The visceral and negative reaction of users—part of what Stark (2016) calls "the emotional context of information privacy"—betrayed the uneven power dynamics that mark our online lives, where dramatic design changes can be foisted on upon millions (or even a billion) users. Recalling Doppelt's discussion of the connection between power and labor, this and other violations by the company points toward one way in which the design of online platforms may be implicated in the development and maintenance of our self-respect.

It is important to point out, however, that privacy violations are not always (or even usually) inflicted equally across all individuals or groups, be they citizens of a nation-state or users of a website. In the United States context, disparities in surveillance across racial and ethnic groups are well established (Parenti 2004; Browne 2015; Bedoya 2016). Today, new surveillance practices stand to further entrench these disparities, as in the case of electronic monitoring for already racially-skewed prison populations (Albert and Delano, 2018). And privacy protections can also undermine human dignity when they are applied unevenly or conceived of inappropriately, as with the uneven privacy protections afforded to seniors in nursing care (Young 2004). As Levy, Kilgour, and Berridge (2019) found in their work on consumer surveillance in elder care facilities, emerging law and policy has tended to defer to residents' family members and legal representatives, leaving little space for the voices of residents and facility employees in deciding how new, lightweight surveillance technologies should be regulated and deployed. Similarly, privacy protections developed to promote liberal ideals of autonomy or dignity in the home can sometimes work to further institutionalize sex- and gender-based power imbalances, reinforcing conditions of domestic confinement, traditional social roles, and violence (Allen 2004, 35).

The issues of privacy, information, and technology implicate Wolff's sources of reduced respect-standing in various ways. Failures of courtesy occur when contextually bound information norms are misunderstood or violated, as when changes to online social networking platforms upend previously established information flows. The widespread deployment of pervasive surveillance technologies against particular racial and ethnic groups can promote an environment of mistrust that systematically targets the dignity and security of particular groups, as exemplified by revelations of domestic spying carried out on Black Lives Matter activists (Vohra 2017; Levin 2018). Finally, the ubiquitous and invasive data-gathering techniques employed online can produce (to use Wolff's term) "revelations" of information, that is, they can unwittingly reveal information, invite undue scrutiny, or have negative social and financial consequences. This risk is especially acute when the vulnerable parties have little say in how information about them is collected or circulated, as with elder care residents and employees. Depending on how these technological practices are employed, they can have the effect of reducing a person's respect-standing—from the upsetting of informational norms to undue subjection to surveillance to forced disclosure.

5.2 Domain 2: Information and Identity

Beyond privacy, the standards and categories imposed by informational technologies can also influence one's sense of self-respect. Information technologies are not neutral or empty vessels for encoding and transmitting information (Briggle and Mitcham 2009, 171)—rather, they necessarily require some more or less complete set of standards, classifications, or protocols in order to function. Without such recognizable and shared standards, advanced communication networks like the Internet would be impossible. In some cases, the standards imposed by these systems are of immediate relevance to a person's sense of self, imposing what Manders-Huits (2010) describes as an information system's "administrative conception" of identity and identification. Importantly, this "administrative" or built-in conception of subjects' identities is, as with the design of online platforms discussed in the previous section, not neutral. These affordances can be discriminatory when they fail to represent certain populations or people, or when they encode assumptions about the world that systematically exclude other ways of understanding phenomena (Wittkower 2018, 22). Today, these problems are amplified by often opaque automated or algorithmic processes (see Cheney-Lippold 2011; Bucher 2018).

For minority or otherwise vulnerable groups, administrative conceptions of personal identity pose a particular threat to self-respect, since these conceptions often come into conflict with our "self-informative" identities (Manders-Huits 2010)—that is, self-conceptions that tend to be more comprehensive, reflexive, and moral in nature. She discusses three ways in which these identities can come into tension. The first, and perhaps most obvious, is the problem of computational reductionism, that is, an "endorsement of the ideal that anything can be expressed in terms of data (and the probabilities and profiles based on them)" (Manders-Huits 2010, 51). Though necessary for the operation of computational systems, practices of computational reductionism cannot take into account "soft information or data, such as contextual and motivational features, background knowledge, and (personal) explanation regarding actions or decisions" (Manders-Huits 2010, 51).

In the US context, the problem of computational reductionism is evident in the practice of body scanning employed by the Transportation Safety Administration (TSA) and the problems it generates for transgender, intersex, and gender non-conforming (GNC) individuals. As Beauchamp (2009), Costanza-Chock (2018) and others have pointed out, the millimeter wave scanning machines employed by the TSA are designed around more or less strict binary (i.e., "male" and "female") assumptions about human bodies that fail to account for the full range of body types and configurations. As Costanza-Chock (2018) summarizes, "anyone whose body doesn't fall within an acceptable range of 'deviance' from a normative binary body type is flagged as 'risky' and subject to a heightened and disproportionate burden of the harms (both small and, potentially, large) of airport security systems and the violence of empire they instantiate" (para. 6). In this case, violations of courtesy, mistrust, and revelation are committed all at once, as trans, intersex, and GNC individuals are often (1) unable to negotiate the categories imposed on them, (2) disproportionately exposed to scrutiny, and (3) routinely forced to reveal information about their bodies, identities, or personal histories that are deemed deviant by the normative standards of the system. Here, social attitudes, institutional structures, and technology design conspire to produce violent conditions (Spade 2015; Hoffmann 2018) hostile to the development and maintenance of these individuals' social bases of self-respect.

Outside of reduction, the persistence of information (particularly digital information online) can shape one's nominal identity in ways that obstruct the development of—or actively harm—one's self-informative identity. Because information captured in files and databases endures, is easily spread, and is often difficult to change or remove, the ability of individuals "to wrest themselves from (former) characterizations and change in light of (new) moral considerations" is stunted (Manders-Huits 2010, 52). Consider, for example, increasingly pervasive forms of online harassment made possible, in part, by the design of online platforms and information systems (Massanari 2017). Online harassment and abuse—which may include threats of violence or physical harm, privacy invasions, defamation, and technical attacks—is more than just a mere extension of offline abuse, as the affordances of networked information systems can accelerate and exacerbate harm or injury (Citron 2014).

In particular, the Internet helps extend the life of destructive or abusive information, making it nearly impossible to forget about or evade harm (Citron 2014, 4). This problem is particularly acute for victims of the ill-named (see Jeong 2015) "revenge porn"—that is, the nonconsensual distribution of sexually graphic images of an individual often (though not always) posted and circulated online with malicious or ill intent (Citron and Franks 2014). These efforts are "inextricably tied to the nature of the Internet" (Levendowski 2014, 426), leveraging its affordances to shame or injure victims (and, subsequently, reduce their respect-standing) in ways that are difficult to remedy and nearly impossible to remove. In this way, the persistence of information online poses an ongoing challenge to victims whose social bases of self-respect have been directly and maliciously targeted.

Lastly, Manders-Huits (2010) draws on Ian Hacking's notion of "dynamic nominalism" to show how moral or self-informative identities often take up or are shaped by available categories, labels, or attributed identifications (52–53). Dynamic nominalism refers to the processes by which a given system watches what you do, fits you into a pattern, then feeds the pattern back to you in the form of options set by the pattern, the options reinforce the pattern, and so on. Importantly, however, these patterns are not solely determined by our individual preferences or behaviors; they are also informed by assumptions in the aggregate and the behavior of others within a system. Safiya Noble (2018) has extensively documented how this dynamic cycle is complicit in reproducing (or even amplifying) racist and sexist cultural ideas—ideas that stand to have the biggest negative impact on those already vulnerable to racism and sexism. For example, she shows how Google searches for the term "black girls" that return results for pornographic web pages reproduce historical conditions of racist, sexualized subjugation for Black women and girls (64–109). As Noble (2018) summarizes, "these search engine results for women whose identities are already maligned in the media, such as Black women and girls, only further debase and erode efforts for social, political, and economic recognition and justice" (88).

Problems of computational reductionism, the persistence of information, and dynamic nominalism can undermine certain individuals' respect-standing according to all three of Wolff's criteria. Information or standards that are imposed on an individual from without—and that endure in ways that are difficult to change—can, as in the case of TSA body scanning practices, produce violations of courtesy and systematic mistrust that systematically undermines the dignity of trans, intersex, and gender non-conforming people. In different ways, online harassment and abuse enabled by online platforms and problematic search results work to shame or degrade particular individuals, especially women. The persistence of online information and processes of dynamic nominalism make these forms of shaming particularly pernicious and often difficult to remedy.

6. Conclusion

As demonstrated by various scholars, our self-respect is informed, in part, by considerations external to the individual. Recalling one of Rawls' (1999) earliest statements on the subject, it is unreasonable to expect that individuals will remain assured of their own value "in the face of enduring contempt or even the indifference of others" (171). While others have shown how institutionalized discrimination within social, economic, or political structures can serve to disempower individuals along racial, gender, sexual, or other lines, I have tried—building on insights from values-conscious design and disability studies—to demonstrate that self-respect is also importantly shaped by the design, dissemination, and use of technology.

Information technology, in particular, plays an important role in codifying, entrenching, and reproducing self-respect's social bases. Issues of privacy and surveillance show how technological advancements threaten individual autonomy and dignity, while uneven patterns of power and surveillance undermine the respect-standing of particular individuals or groups. Additionally, the collection, classification, and implementation of information pose a distinct set of threats stemming from practices of computational reductionism, the persistence of information, and processes of dynamic nominalism (Manders-Huits 2010). Biased, discriminatory, or incomplete standards, especially when deployed on a massive scale, can serve to systematically undermine the dignity of certain individuals or groups, while the persistence of online information can work to shame or degrade in pernicious ways. When coupled with self-respect's social dimensions, the values and affordances embedded in the design and use of information technology plays a key role in promoting the development of self-respect for some people and hindering it for others. In view of this, work interested in the practical relationship between information, technology, and social justice ought to be mindful of the importance of self-respect and its sociotechnical bases.

References

Albert, Kendra, and Maggie Delano. 2018. "The World Is a Prison." *Logic Magazine*, April 1, 2018. https://logicmag.io/03-the-world-is-a-prison/.

Allen, Anita L. 2004. "Privacy in American Law." In *Privacies: Philosophical Evaluations*, edited by Beate Rössler, 19–39. Stanford, CA: Stanford University Press.

Barnes, Colin. 2012. "Understanding the Social Model of Disability: Past, Present, and Future." In *Routledge Handbook of Disability Studies*, edited by Nick Watson, Alan Roulstone, and Carol Thomas, 12–29. London: Routledge.

Beauchamp, Toby. 2009. "Artful Concealment and Strategic Visibility: Transgender Bodies and U.S. State Surveillance After 9/11." *Surveillance & Society* 6, no. 4: 356–66.

Bedoya, Alvaro M. 2016. "The Color of Surveillance." *Slate*, January 18, 2016. http://www.slate.com/articles/technology/future_tense/2016/01/what_the_fbi_s_surveillance_of_martin_luther_king_says_about_modern_spying.html.

Bell, Chris. 2010. "Is Disability Studies Actually White Disability Studies?" In *The Disability Studies Reader*, edited by Lennard J. Davis, 3rd ed., 266–73. New York, NY: Routledge.

Benn, Stanley I. 1971. "Privacy, Freedom, and Respect for Persons." In *Nomos XIII: Privacy*, edited by J. Roland Pennock and John W. Chapman, 1–27. New York, NY: Atherton.

Bloustein, Edward J. 1984. "Privacy as an aspect of human dignity: An answer to Dean Prosser." In *Philosophical Dimensions of Privacy: An Anthology*, edited by Ferdinand D. Schoeman, 156–202. Cambridge, UK: Cambridge University Press.

Boxill, Bernard. 1992. "Two Traditions in African American Political Philosophy." *Philosophical Forum* 24, no. 1–3: 119–35.

Boxill, Bernard R. 1976. "Self-Respect and Protest." *Philosophy & Public Affairs* 6, no. 1: 58–69.

Braman, Sandra. 2006. *Change of State: Information, Policy, and Power*. Cambridge, MA: MIT Press.

Brey, Philip. 2007. "Theorizing the Cultural Quality of New Media." *Techné: Research in Philosophy and Technology* 11, no. 1: 2–18.

Brey, Philip, and Luciano Floridi. 2010. "Values in Technology and Disclosive Computer Ethics." In *The Cambridge Handbook of Information and Computer Ethics*, 41–58. Cambridge, UK: Cambridge University Press.

Briggle, Adam, and Carl Mitcham. 2009. "From the Philosophy of Information to the Philosophy of Information Culture." *The Information Society* 25, no. 3: 169–174. doi:10.1080/01972240902848765.

Browne, Simone. 2015. *Dark Matters: On the Surveillance of Blackness*. Durham, NC: Duke University Press.

Bucher, Taina. 2018. *If . . . Then: Algorithmic Power and Politics*. Oxford and New York: Oxford University Press.

Charles, Sonya. 2010. "How Should Feminist Autonomy Theorists Respond to the Problem of Internalized Oppression?" *Social Theory & Practice* 36, no. 3: 409.

Cheney-Lippold, John. 2011. "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control." *Theory, Culture & Society* 28, no. 6: 164–181. doi:10.1177/0263276411424420.

Citron, Danielle Keats. 2014. *Hate Crimes in Cyberspace*. Harvard University Press.

Citron, Danielle Keats, and Mary Anne Franks. 2014. "Criminalizing Revenge Porn." *Wake Forest Law Review* 49, no. 2: 345–391.

Cohen, Joshua. 2003. "For a Democratic Society." In *The Cambridge Companion to Rawls*, edited by Samuel Freeman, 86–138. Cambridge, UK: Cambridge University Press.

Cohen, Julie E. 2012. *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*. New Haven, CT: Yale University Press.

Costanza-Chock, Sasha. 2018. "Design Justice, A.I., and Escape from the Matrix of Domination." *Journal of Design and Science*, July. doi:10.21428/96c8d426.

Darwall, Stephen L. 1977. "Two Kinds of Respect." *Ethics* 88, no. 1: 36–49.

Dillon, Robin S. 1997. "Self-Respect: Moral, Emotional, Political." *Ethics* 107, no. 2: 226.

Dillon, Robin S. 2010. "Respect for Persons, Identity, and Information Technology." *Ethics and Information Technology* 12, no. 1: 17–28. doi:10.1007/s10676-009-9188-8.

Docherty, Daniel, Richard Hughes, Patricia Phillips, David Corbett, Brendan Regan, Andrew Barber, Michael Adams, Kathy Boxall, Ian Kaplan, and Shayma Izzidien. 2010. "This Is What We Think." In *The Disability Studies Reader*, edited by Lennard J. Davis, 3rd ed., 432–440. New York, NY: Routledge.

Dodgson, Lindsay. 2018. "Men Are Getting the Credit for Women's Work through Something Called 'Hepeating' — Here's What It Means." *Business Insider*, March 8, 2018. https://www.businessinsider.com/what-is-hepeating-2017-9.

Doppelt, Gerald. 1981. "Rawls' System of Justice: A Critique from the Left." *Nous* 15, no. 3: 259–307.

Doppelt, Gerald. 2009. "The Place of Self-Respect in a Theory of Justice." *Inquiry* 52, no. 2: 127–154. doi:10.1080/00201740902790219.

Eyal, Nir. 2005. "'Perhaps the Most Important Primary Good': Self-Respect and Rawls's Principles of Justice." *Politics, Philosophy & Economics* 4, no. 2: 195–219. doi:10.1177/1470594X05052538.

Flanagan, Mary, Daniel C. Howe, and Helen Nissenbaum. 2008. "Embodying Values in Technology: Theory and Practice." In *Information Technology and Moral Philosophy*, edited by Jeroen van den Hoven and John Weckert, 322–353. Cambridge, UK: Cambridge University Press.

Friedman, Batya, Peter H. Kahn, Alan Borning, Ping Zhang, and Dennis F. Galleta. 2006. "Value Sensitive Design and Information Systems." In *Human-Computer Interaction and Management Information Systems: Foundations*, 348–372. Armonk, NY: M. E. Sharpe.

Friedman, Batya, and Helen Nissenbaum. 1996. "Bias in Computer Systems." *ACM Transactions on Information Systems* 14, no. 3: 330–347.

Garland-Thomson, Rosemarie. 2006. "Ways of Staring." *Journal of Visual Culture* 5, no. 2: 173–192. doi:10.1177/1470412906066907.

Gillespie, Tarleton. 2010. "The Politics of 'Platforms.'" *New Media & Society* 12, no. 3: 347–364. doi:10.1177/1461444809342738.

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*. New Haven, CT: Yale University Press.

Gugliucci, Nicole. 2017. "My Friends Coined a Word: Hepeated. For When a Woman Suggests an Idea and It's Ignored, but Then a Guy Says Same Thing and Everyone Loves It." Tweet. *@NoisyAstronomer* (blog). September 22, 2017. https://twitter.com/NoisyAstronomer/status/911213826527436800?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E911213826527436800&ref_url=http%3A%2F%2Fuk.businessinsider.com%2Fwhat-is-hepeating-2017-9.

Hoffmann, Anna Lauren. 2017. "Beyond Distributions and Primary Goods: Assessing Applications of Rawls in Information Science and Technology Literature since 1990." *Journal of the Association for Information Science and Technology* 68, no. 7: 1601–1618. doi:10.1002/asi.23747.

Hoffmann, Anna Lauren. 2018. "Data, Technology, and Gender: Thinking about (and from) Trans Lives." In *Spaces for the Future: A Companion to Philosophy of Technology*, edited by Joseph C. Pitt and Ashley Shew, 3–13. New York, NY: Routledge.

Jeong, Sarah. 2015. "Snap Judgment: In Between Revenge Porn and Sex Work." *Bitch Media*, December 2, 2015. https://www.bitchmedia.org/article/snap-judgment-revenge-porn.

Kline, Stephen J. 2003. "What Is technology?" In *Philosophy of Technology: The Technological Condition*, edited by Robert C. Scharff and Val Dusek, 386–397. Malden, MA: Blackwell. (Reprinted from *Bulletin of Science, Technology & Society, 1*: 215–218, 1980).

Klonick, Kate. 2017. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review* 131: 1598–1670.

Levendowski, Amanda. 2014. "Using Copyright to Combat Revenge Porn." *New York University Journal of Intellectual Property and Entertainment* 3, no. 2: 422–446.

Levin, Sam. 2018. "Black Activist Jailed for His Facebook Posts Speaks out about Secret FBI Surveillance." *The Guardian*, May 11, 2018, sec. World news. https://www.theguardian.com/world/2018/may/11/rakem-balogun-interview-black-identity-extremists-fbi-surveillance.

Levy, Karen, Lauren Kilgour, and Clara Berridge. 2019. "Regulating Privacy in Public/Private Space: The Case of Nursing Home Monitoring Laws." *Elder Law Journal* 26: 323–363.

Manders-Huits, Noëmi. 2010. "Practical versus Moral Identities in Identity Management." *Ethics and Information Technology* 12: 43–55.

Manders-Huits, Noëmi, and Michael Zimmer. 2009. "Values and Pragmatic Action: The Challenges of Introducing Ethical Intelligence in Technical Design Communities." *International Review of Information Ethics* 10, no. 2: 37–44.

Marx, Karl. 1975. "Economic and Philosophical Manuscripts." In *Early Writings.* Translated by G. Benton, 279–400). London, UK: Penguin. (Original work published in 1844.)

Massanari, Adrienne. 2017. "#Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures." *New Media & Society* 19, no. 3: 329–346. doi:10.1177/1461444815608807.

Mathiesen, Kay. 2015. "Toward a Political Philosophy of Information." *Library Trends* 63, no. 3: 427–647. doi:10.1353/lib.2015.0000.

Mohr, Richard. 1988. *Gays/Justice: A Study of Ethics, Society, and Law*. New York, NY: Columbia University Press.

Moody-Adams, Michelle. 1993. “Race, Class, and the Social Construction of Self-Respect.” *Philosophical Forum* 24, no. 1–3: 251–266.

Nissenbaum, Helen. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, 1st ed. Stanford, CA: Stanford Law Books.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Nussbaum, Martha. 2004. “The Future of Feminist Liberalism.” In *Varieties of Feminist Liberalism*, edited by Amy Baehr, 103–132. Lanham, MD: Rowman and Littlefield.

Okin, Susan Moller, and Amy Baehr. 2004. “Is Multiculturalism Bad for Women?” In *Varieties of Feminist Liberalism*, 191–206. Lanham, MD: Rowman and Littlefield.

Oliver, Michael. 1981. “A New Model of the Social Work Role in Relation to Disability.” In *The Handicapped Person*, edited by Jo Campling, 19–32. London, UK: Royal Association for Disability Rights.

Parenti, Christian. 2004. *The Soft Cage: Surveillance in America from Slavery to the War on Terror,* reprint ed. New York, NY: Basic Books.

Parker, Morgan. 2017. “How to Stay Sane While Black.” *The New York Times*, December 22, 2017, sec. Opinion. https://www.nytimes.com/2016/11/20/opinion/sunday/how-to-stay-sane-while-black.html.

Pilapil, Renante D. 2014. “Recognition as Redistribution.” *Critical Horizons* 15, no. 3: 284–305. doi:10.1179/1440991714Z.00000000036.

Rawls, John. 1971. *A Theory of Justice,* rev. ed. Cambridge, MA: Belknap Press.

Rawls, John. 1993. *Political Liberalism*. New York: Columbia University Press.

Rawls, John. 1999. “Distributive Justice: Some Addenda.” In *Collected Papers*, edited by Samuel Freeman, 154–175. Cambridge, MA: Harvard University Press. (Reprinted from Rawls, 1968. “Distributive Justice: Some Addenda,” *The American Journal of Jurisprudence* 13, no. 1: 51–71.)

Rawls, John. 2007. “Lectures on Marx.” In *Lectures on the History of Political Philosophy*, edited by Samuel Freeman, 319–74. Cambridge, MA: Belknap Press.

Regan, Priscilla M. 1995. *Legislating Privacy: Technology, Social Values, and Public Policy*. Chapel Hill: University of North Carolina Press. https://www.jstor.org/stable/10.5149/9780807864050_regan.

Reiman, Jeffrey H. 1976. “Privacy, Intimacy, and Personhood.” *Philosophy & Public Affairs* 6, no. 1: 26–44.

Roberts, Sarah T. 2018. “Digital Detritus: ‘Error’ and the Logic of Opacity in Social Media Content Moderation.” *First Monday* 23, no. 3. http://firstmonday.org/ojs/index.php/fm/article/view/8283.

Schoeman, Ferdinand D. 1984. “Privacy and Intimate Information.” In *Philosophical Dimensions of Privacy: An Anthology*, edited by Ferdinand D. Schoeman, 203–418. Cambridge, UK: Cambridge University Press.

Shakespeare, Tom. 2010. “The Social Model of Disability.” In *The Disability Studies Reader*, edited by Lennard J. Davis, 3rd ed., 266–273. New York, NY: Routledge.

Shew, Ashley. 2017. “Technoableism, Cyborg Bodies, and Mars.” *Technology and Disability* (blog). November 11, 2017. https://techanddisability.com/2017/11/11/technoableism-cyborg-bodies-and-mars/.

Solove, Daniel J. 2010. *Understanding Privacy,* Feb. 28, 2010 ed. Cambridge, MA and London, UK: Harvard University Press.

Spade, Dean. 2015. *Normal Life: Administrative Violence, Critical Trans Politics, and the Limits of Law*. Durham, NC: Duke University Press.

Stark, Luke. 2016. “The Emotional Context of Information Privacy.” *The Information Society* 32, no. 1: 14–27. doi:10.1080/01972243.2015.1107167.

Suzor, Nicolas. 2018. "Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms." *Social Media + Society* 4, no. 3: 2056305118787812. doi:10.1177/2056305118787812.

Telfer, Elizabeth. 1968. "Self-Respect." *The Philosophical Quarterly (1950–)* 18, no. 71: 114–121. doi:10.2307/2217509.

Terzi, Lorella. 2010. "What Metric of Justice for Disabled People? Capabiity and Disability." In *Measuring Justice: Primary Goods and Capabilities*, edited by Harry Brighouse and Ingrid Robeyns, 150–173. Cambridge, UK: Cambridge University Press.

Thomas, L. 1995. "Self-Respect: Theory and Practice." In *Dignity, Character, and Self-Respect*, edited by R. S. Dillon, 251–270. London, UK: Routledge. (Reprinted from Thomas, L. 1983. "Self-Respect: Theory and Practice," In *Philosophy Born of Struggle: Anthology of Afro-American Philosophy from 1917*, edited by L. Harris, 174–189. Dubuque, IA: Kendall/Hunt).

Tufekci, Zeynep. 2015. "Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency." *Colorado Technology Law Journal* 13: 203–17.

UPIAS. 1976. *Fundamental Principles of Disability*. London, UK: Union of the Physically Impaired Against Segregation.

Vaidhyanathan, Siva. 2018. *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford, UK: Oxford University Press.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford, UK: Oxford University Press.

Verbeek, Peter-Paul. 2009. "Moralizing Technology: On the Morality of Technological Artifacts and Their Design." In *Readings in the Philosophy of Technology*, edited by David M. Kaplan, 2nd ed., 226–43. Lanham, MD: Rowman and Littlefield.

Vohra, Sweta. 2017. "Documents Show US Monitoring of Black Lives Matter." *Al Jazeera*, November 28, 2017. https://www.aljazeera.com/news/2017/11/documents-show-monitoring-black-lives-matter-171128110538134.html.

Warren, Samuel D., and Louis D. Brandeis. 1890. "The Right to Privacy." *Harvard Law Review* 4, no. 5: 193–220. doi:10.2307/1321160.

Westin, Alan F. 1967. *Privacy and Freedom*. New York, NY: Atheneum.

Wittkower, D. E. 2016. "Principles of Anti-Discriminatory Design." In *2016 IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS)*, 1–7. doi:10.1109/ETHICS.2016.7560055.

Wittkower, D. E. 2018. "Discrimination." In *Spaces for the Future: A Companion to Philosophy of Technology*, edited by Joseph C. Pitt and Ashley Shew, 14–28. New York, NY: Routledge.

Wolff, Jonathan. 1998. "Fairness, Respect, and the Egalitarian Ethos." *Philosophy & Public Affairs* 27, no. 2: 97–122. doi:10.1111/j.1088–4963.1998.tb00063.x.

Young, Iris Marion. 1990. *Justice and the Politics of Difference*. Paperback reissue. Princeton, N.J: Princeton University Press.

Young, Iris M. 2004. "A Room of One's Own: Old Age, Extended Care, and Privacy." In *Privacies: Philosophical Evaluations*, edited by Beate Rössler, 168–186. Stanford, CA: Stanford University Press.

Young, Iris Marion. 2006. "Taking the Basic Structure Seriously." *Perspectives on Politics* 4, no. 1: 91–97. doi:10.1017/S1537592706060099.

Zimmer, Michael, and Anna Lauren Hoffmann. 2011. "Privacy, Context, and Oversharing: Reputational Challenges in a Web 2.0 World." In *The Reputation Society: How Online Opinions Are Reshaping the Offline World*, edited by Hassan Massum and Mark Tovey, 175–184. Cambridge, MA: MIT Press.

Zink, James R. 2011. "Reconsidering the Role of Self-Respect in Rawls's A Theory of Justice." *The Journal of Politics* 73, no. 2: 331–44. doi:10.1017/S0022381611000302.

CHAPTER 13

FREEDOM IN AN AGE OF ALGOCRACY

JOHN DANAHER

1. Introduction

We live in an age of algorithmic governance. Advances in computing technology have created a technological infrastructure that permeates, shapes, and mediates our everyday lives. From personal computers to smartphones, from high street banks to high school playgrounds, from dawn to dusk, we are continually monitored, nudged, and reminded by a growing network of "smart" devices. The native language of these devices is that of the algorithm. We are nestled inside a web of them: algorithms that collect, parse, and sort our data; algorithms that spot patterns and learn from their mistakes; algorithms that issue instructions to the world.

Scholars in many social science disciplines are now trying to map, analyze, and evaluate the consequences of this rise in algorithmic governance. They give it different names, depending on their disciplinary backgrounds and scholarly interests. Some favor terms such as "algorithmic regulation" (Yeung 2017; 2018) or "algorithmic governmentality" (Rouvroy 2013x; 2015). I favor using the neologism "algocracy," first coined by the sociologist A. Aneesh (2006; 2009), to describe the phenomenon (Danaher 2016a; 2016b). The label itself is not as important as the phenomenon it is used to describe: the unavoidable and seemingly ubiquitous use of computer-coded algorithms to understand and control the world in which we live.

Within the growing scholarly literature on algocracy, three major debates have emerged. The first is the debate about *privacy* and *surveillance* (Polonetsky and Tene 2013). Contemporary algorithmic governance is made possible through the use of Big Data systems. These systems are what drive machine learning algorithms to develop impressive abilities to sort and mine data. These systems would not be possible without mass surveillance, which always poses a threat to privacy. The second is the debate about *bias* and *inequality* (Zarsky 2012; Crawford 2013; O'Neil 2016; Binns 2018; Eubanks 2018; Noble 2018). The machine learning algorithms that are used to develop (among other things) credit scores for bank customers, or to predict likely rates of recidivism among prisoners, are created through the use of "training" data. That is, they learn how to predict future behaviors by spotting patterns in large databanks of past behaviors. There is considerable concern that these databanks, and the patterns that are extrapolated from them, can be biased. Studies have been done trying to highlight the biases that arise in different domains, such as the apparent racial bias of the predictive policing and predictive sentencing algorithms that are widely used in the United States (Ferguson 2017; Binns 2018). The third is the debate about *transparency* and *procedure*. A major concern about machine learning algorithms is that they are relatively opaque in how they operate (Pasquale 2015; Danaher 2016b). These algorithms can produce impressive results, but their precise workings are often hidden from view, for both legal and technical reasons (Pasquale 2015; Burrell 2016). They are "black boxes": changing the world around them without being readily comprehensible to the human beings affected. This threatens many of the procedural values that we hold dear in democratic societies (Citron and Pasquale 2014; Danaher 2016b).

Slightly less widely discussed, though every bit as important (Yeung 2017; Frischmann and Selinger 2018), is the impact that algocratic systems have on our freedom. With all these systems monitoring, nudging, and prompting, obvious questions arise about the effects this has on individual choice and behavior. Do algocratic systems promote or undermine our freedom? How should we respond to this? Some people have started to address these questions, but those that do tend to fixate on the *threats* that algocratic systems pose to our freedom (Yeung 2017; Frischmann and Selinger 2018). This can seem understandable: Anything that monitors, nudges, and suggests must, surely, be corroding our freedom? But in this chapter I argue that this is not necessarily the case. To be more precise, I argue that we should resist any seemingly simple answer to the question of whether algocracy negatively or positively impacts our freedom. Freedom is a complex and multidimensional value, and algocracy is a complex and multidimensional phenomenon. It follows, reasonably enough, that the impact of this technological phenomenon on freedom is multifaceted. It can promote and undermine our freedom at one and the same time. We need to be sensitive to this complexity. Only then will we know what it means to be free in an age of algocracy.

To make this case, the chapter proceeds in three main steps. First, it discusses the nature of freedom, illustrating how it is complex in two distinct, but equally significant ways. Second, it discusses the nature of algocracy, explaining the different forms it can take, and suggesting that there is a large "logical space" of different possible algocratic systems in any given domain. Third, taking this complexity onboard, it presents five mechanisms through which algocratic systems can promote and undermine freedom. In each section, the intention is not so much to reach definitive normative conclusions about the desirability/undesirability of algocratic systems; rather, the intention is to present a series of frameworks for thinking about this issue.

Consequently, the chapter is programmatic in nature, favoring breadth of analysis over depth of argumentation.

2. The Complexity of Freedom and How to Think About It

To assess the impact of algocracy on freedom we must have some sense of what freedom consists in. There is a long-standing philosophical debate about what we might call "metaphysical freedom." This debate focuses on the question of whether humans have free will, and the related question of whether human behavior is causally determined and if that prevents us from having free will. This can be contrasted with the debate about "political" or "social" freedom, which focuses less on the deep metaphysical questions, and more on what it takes to live freely within a particular political arrangement, society or culture. Can you be free if the government imposes sanctions on you for following your conscience on religious matters? Can you be free if your workplace, college or school has a speech code that prevents you from saying and doing certain things? These kinds of questions are central to the political tradition of liberalism (broadly conceived).

Although both debates are important, I focus on the political and social form of freedom in what follows. The primary reason for this is that I think the positions staked out in the metaphysical debate are largely unaffected by technological and social change. Whether we have free will or not depends on deep (possibly unknowable) structural features of our reality. It is not something that is going to be changed or affected by the development of a new technology like an algocratic system. Furthermore, the metaphysical and political debates already work largely independently of one another in the philosophical literature. That said, it would be foolish to deny the fact that there is some overlap between the debates. In particular, it is important to realize that those who espouse a compatibilist view of free will—that is, a view which holds that metaphysical freedom is compatible with causal determinism—often develop accounts of what it means to be free that focus on conditions similar to the political ones discussed later in the chapter (e.g., Frischmann and Selinger 2018). Nevertheless, I will not be directly engaging with what compatibilists have to say in the remainder of this chapter.

Even if we limit the focus to the political form of freedom, there is still much that needs to be clarified if we are going to assess the impact of algocracy on it. This is because political freedom is complex in two distinct ways: (1) it is complex with respect to the conditions that need to be satisfied in order to protect it; and (2) it is complex with respect to the way in which it is valued. Let's consider both forms of complexity in more detail.[1]

With respect to the first form of complexity, it is obvious that there are many conceptions of freedom out there and that within these conceptions different conditions are identified as being freedom-promoting or freedom-undermining. It is not hard to find examples of this. The intellectual historian, Quentin Skinner, for instance, has mapped out a genealogy of all the different conceptions of freedom that have been defended since the birth of modern "liberal" political philosophy in the seventeenth century (Skinner 2008a, 2008b, and 2012). He argues that this genealogy has three main branches to it. First, there are those who insist that to be free means that you are free of *interference*, where interference can consist of the use of physical force or coercive threats to direct behavior. Second, there are those who insist that to be free means that you act in a way that is consistent with your *authentic* self, where this involves some consistency between action and your personal values or intrinsic nature. And third, there are those that insist that to be free means to be free from *domination* (Pettit 2001, 2011 & 2014), where domination arises whenever one person's actions are subject to the arbitrary will of another (e.g., they require tacit or explicit approval from that person in order to live an unencumbered life). The first and third branches correspond to the idea of "negative" liberty; the second corresponds to the idea of "positive" liberty.

The philosophers Christian List and Laura Vallentini (2016) take a more abstract approach. They argue that if you survey the literature on political freedom, it is possible to discern from this a "logical space" in which the various conceptions of freedom arise. This might sound like a daunting idea, but it is quite straightforward. You can construct a logical space by identifying the different dimensions along which theories vary. List and Vallentini argue that there are two such dimensions in the case of freedom. They argue that most theorists agree that interference by other agents is something that undermines freedom, but they then disagree on two things: (1) whether that interference is freedom-undermining only if it exists in the actual world or if it can be freedom-undermining if it exists in other possible worlds that are similar to our own (this defines the "modal dimension" of the logical space); and (2) whether some forms of interference should not be taken to undermine freedom because they are morally justified or whether moral and immoral forms of interference should both be taken to undermine freedom (the "moral dimension" of the logical space). Using these two dimensions, List and Vallentini construct a two-by-two matrix that defines four logically possible concepts of freedom. Two of them (those that focus on interferences in the actual world only) correspond to classical liberal theories of freedom as non-interference, similar to those discussed by Skinner. One of them (the one that focuses on the absence of immoral interference across several possible worlds) corresponds to the theory of freedom as non-domination, which is favored by Skinner and Philip Pettit.[2] Finally, there is something that List and Vallentini call the theory of freedom as independence (which involves the absence of moral and immoral interference across several possible worlds), which they argue has been neglected in the debate thus far.

One more example of the complexity of freedom can be found in discussions of autonomous decision making. The relationship between freedom and autonomy is, itself, somewhat complex, with one popular account holding that the former is a "local" property that applies to specific choices (was this choice free?) whereas the latter is a "global" property that applies across someone's lifetime (is this person living an autonomous life?) (Dworkin 1988). For present purposes I will treat them as equivalent concepts. The important point for now is that within the debate about autonomy there are several conditions that need to be satisfied in order for a choice (or a life) to count as autonomous. Everyone agrees on the basic idea—to be autonomous means that you are, in some sense, the "author" of your own life/choices—but there are differences when it comes to the critical sub-conditions of autonomy (Killmister 2017). To give but one example of this, consider the theory of autonomy that was first proposed by Joseph Raz back in the 1980s. This theory focuses on three conditions that need to be satisfied if a particular choice is to count as autonomous:

> If a person is to be maker or author of his own life then he must have the mental abilities to form intentions of a sufficiently complex kind, and plan their execution. These include minimum rationality, the ability to comprehend the means required to realize his goals, the mental faculties necessary to plan actions, etc. For a person to enjoy an autonomous life he must actually use these faculties to choose what life to have. There must in other words be adequate options available for him to choose from. Finally, his choice must be free from coercion and manipulation by others, he must be independent.
>
> (Raz 1986, 373)

The three conditions of autonomy embedded in this quoted passage are (1) the person must have the *minimum rationality* to plan actions that will allow them to achieve their goals; (2) they must have *adequate options* available to choose from; and (3) they must be *independent*, which Raz takes to mean free from coercion and manipulation when making and implementing their choices.

I could go on, but I think the point is made. Theories of freedom are complex. There are different accounts of what it means to be free and within those different accounts many distinct freedom-undermining and promoting conditions have been identified. It is tempting at this point to deal with this complexity by throwing your hands up in despair and simply picking and defending one preferred theory. That is what many people do. But I believe that this is mistaken. There is much less tension and disagreement between the different accounts than first appears to be the case. Indeed, any apparent tension can be largely dissolved by acknowledging that all the different accounts identify some conditions that are relevant to freedom, but that the accounts vary in the breadth of the conditions they deem to be important or in how much weight they place on them. It is possible to accommodate this variety by viewing freedom as a scalar property and not a binary one. In other words, by accepting that people can be more or less free, and not simply free or un-free. Assessing the scale of freedom that any one individual has will then depend on the various conditions considered above (rationality, adequate options, absence of interference and/or domination).

This doesn't mean, however, that we should view freedom as something that varies along a single dimension and that degrees of freedom can be easily determined by one's location along that single dimension. As should be clear from the preceding discussion, the different conditions of freedom identified by the different theories of freedom suggest that the reality is far more complex. It is more likely that the conditions define a multi-dimensional space and that one's degree of freedom depends on where one fits within that multi-dimensional space. Models of this multi-dimensional space could be quite conceptually unwieldy, depending on whether you are a "lumper" or "splitter" when it comes to defining the different dimensions. For ease of understanding, and for illustrative purposes, I will adopt a three-dimensional model of freedom over the remainder of this chapter. This model focuses on the following three dimensions of freedom:

1. *The Intelligibility/Rationality Dimension:* which measures the extent to which one can make decisions that are based on identifying, weighing, and assessing options for their fit with one's preferences and plans. This does not focus on some narrow form of "economic" rationality; it focuses on the ability to make decisions in an intelligible fashion.
2. *The Manipulation Dimension:* which measures the extent to which one's decisions are free from manipulation, where manipulation can come in the form of physical force, mental coercion, brainwashing, cultural indoctrination and so on. Some of these are classed as highly manipulative, and some less so.
3. *The Domination Dimension*: which measures the extent to which one's decisions are free from domination, where domination involves the presence of some authorizing agent from whom one must tacitly or explicitly get approval in order to act.

You could probably parse the dimensional space in different ways, but I think each of these is a defensible inclusion within a complex account of freedom. I also think that if we understand them to define distinct dimensions of freedom, we can appreciate something important: the possible need for tradeoffs across the different dimensions. It may turn out, for example, that completely avoiding all forms of manipulation will require that we sacrifice some degree of intelligibility, or that removing some forms of manipulation requires accepting some forms of domination. In other words, it may be impossible to maximize along all dimensions simultaneously. We may find out that we have to prioritize or compromise when it comes to protecting our freedom.

This is still only the first form of complexity we must confront when thinking about freedom—the complexity of the conditions/dimensions of freedom. We still have to confront the complexity with respect to how freedom is valued. It is all well and good to have a clear sense of what freedom requires, but this is useless if we don't know why we are so obsessed with it in the first place. Fortunately, there is less complexity to contend with here. There are essentially three different approaches we can take to the value of freedom. We can view freedom as an intrinsic value; that is, something worth protecting and promoting in and of itself. We can view it as an instrumental value; that is, something worth protecting because it helps us to achieve other valuable ends like well-being or flourishing. Or we can view it as a bit of both; that is, as something that is both intrinsically valuable and instrumentally valuable.

There are some approaches to the value of freedom that sit outside this simple tripartite scheme. For instance, Ian Carter (1995 and 1999) has defended the view that freedom is an "independent" value, which he defines as being slightly distinct from an intrinsic value. Likewise, I have argued that freedom should be viewed as an *axiological catalyst*; that is. as something that makes good things better and bad things worse (Danaher

2018). In saying this, I was motivated by the fact that a good deed freely done is usually judged more favorably than a good deed performed under coercion; and someone who killed a bunch of people freely is usually judged less favorably than someone who did so unfreely. It is also possible to think that freedom is completely devoid of value and shouldn't be protected at all.

The subtleties of these other positions lie beyond the scope of this paper, but even if we stick with the tripartite view there is still plenty of complexity that needs to be worked out. If we value freedom intrinsically, then we will need to decide where freedom fits within the pantheon of other values like friendship, knowledge, pleasure, flourishing and so on. Is freedom the single most important value? Does it rank equally among these other values? Or is it less important? Similarly, if we value freedom instrumentally, then we need to determine how important a means to other valuable ends it is. Could it be that there are other, more efficient, ways to achieve these ends? Or is freedom the single most reliable means to these ends? How we answer these questions will determine, in large part, our attitude toward a phenomenon like algocracy. If we think freedom is the single most important value, then we might view any threat to it as a major social problem that needs to be addressed with utmost speed. If we think it is just one value among many, and possibly not that important, we might be willing to sacrifice some degree of freedom to protect these other ends.

I won't say much about how we should or might value freedom in what follows. I will focus, instead, on the question of how algocracy might affect the various dimensions of freedom. But, clearly, the weight of the arguments I make, and the attitude you should take toward them, will depend significantly on how value freedom.

3. The Logical Space of Algocracy

The complexity of freedom is just one side of the coin. We also have to consider the complexity of algocracy. To do this, we first need to have a clear sense of what algocracy is. I said at the start that "algocracy" is my preferred term for an increasingly familiar phenomenon: the use of big data, predictive analytics, machine learning, AI, robotics (etc.) in any system that governs human behavior. The term was originally coined by the sociologist A. Aneesh (2006; 2008). Aneesh's main interest was in delineating between the different forms that human governance systems can take. A governance system can be defined, roughly, like this:

> **Governance system**: Any system that structures, constrains, incentivizes, nudges, manipulates or encourages different types of human behavior.

This is a very broad definition, but this is deliberate since "governance" is taken to be a broad concept. It's natural to speak of governance as something that arises at an institutional or governmental level, and that is certainly an obvious home for the concept, but it is also something that arises outside of a formal institutional context (e.g., governance by tacit social norms) and at an individual level (what tools do I use to govern my own behavior). Aneesh drew a contrast between three main types of governance system in his research: markets, bureaucracies and algocracies. A market is a governance system in which prices structure, constrain, incentivize, nudge (etc.) human behavior; a bureaucracy is a governance system in which rules and regulations structure, constrain, incentivize, nudge (etc.) human behavior; and an algocracy is:

> **Algocracy**: A governance system in which computer coded algorithms structure, constrain, incentivize, nudge, manipulate or encourage different types of human behavior.[3]

Aneesh used the concept to understand how workers participated in a globalized economy. Aneesh thought it was interesting how more workers in the developing world were working for companies and organizations that were legally situated in other jurisdictions. He argued that this was due to new technologies (computers + internet) that facilitated remote work. This gave rise to new algocratic governance systems within corporations, which sidestepped or complemented the traditional market or bureaucratic governance systems within such organizations.

That's the origin of the term. I tend to use the term in a related but slightly different sense. I certainly look on algocracies as kinds of governance system—ones in which behavior is shaped by algorithmically programmed architectures. But I also use the term by analogy with "democracy," "aristocracy," and "technocracy." In each of those cases, the suffix "cracy" is used to mean "rule by" and the prefix identifies whoever does the ruling. So "democracy" is "rule by the people" (the *demos*), aristocracy is "rule by aristocrats" and so on. Algocracy then can also be taken to mean "rule by algorithm," with the emphasis being on rule. In other words, for me "algocracy" captures the authority that is given to algorithmically coded architectures in contemporary life. Whenever you are denied a loan by a credit-scoring algorithm; whenever you are told which way to drive by a GPS routing-algorithm; whenever you are prompted to exercise a certain way or eat a certain food by a health and fitness app you are living within an algocratic system.

With this understanding in place, we can already begin to see that algocracy is a complex phenomenon. Algocratic systems arise in different domains (financial, legal, bureaucratic. personal) and take different forms. There have been several attempts to bring order to this complexity. One method of doing so is to focus on the various stages involved in the construction and implementation of an algocratic system. Algocratic systems do things: they make recommendations; they set incentives; they structure possible forms of behavior; and so on. How do they manage this? Much of the answer lies how they use data. Zarsky (2013) suggests that there are three main stages in this: (1) a data collection stage (where information about the world and relevant human beings is collected and fed into the system); (2) a data analysis stage (where algorithms structure, process and organize that data into useful or salient chunks of information); and (3) a data usage stage (where the algorithms make recommendations or decisions based on the information they have processed). Citron and Pasquale (2014) develop a similar framework, using slightly different terminology, that focuses on four main stages. This is illustrated in Figure 13.1.

Effectively, what they do is break Zarsky's "usage" stage into two separate stages: a dissemination stage (where the information processed and analyzed by the algorithms gets communicated to a decision maker) and a decision-making stage (where the decision maker uses the information to do something concrete to an affected party, e.g., deny them a loan because of a bad credit score).

In doing this Citron and Pasquale make an interesting assumption about how the algocratic system relates to the human beings who are affected by it. They assume that the primary function of an algocratic system is to generate recommendations to humans, who still retain ultimate decision-making authority. But this may not be the case. Indeed, as they themselves note, there are different ways in which an algocratic system could connect with (or avoid connecting with) the humans whose behavior is being governed. Adopting a simple tripartite framework originally developed in the military context, they distinguish between human-in-the-loop systems (where humans retain ultimate decision-making authority), human-on-the-loop systems (where humans retain veto power over the algocratic system) and human-off-the-loop systems (where the system functions without human input or oversight).

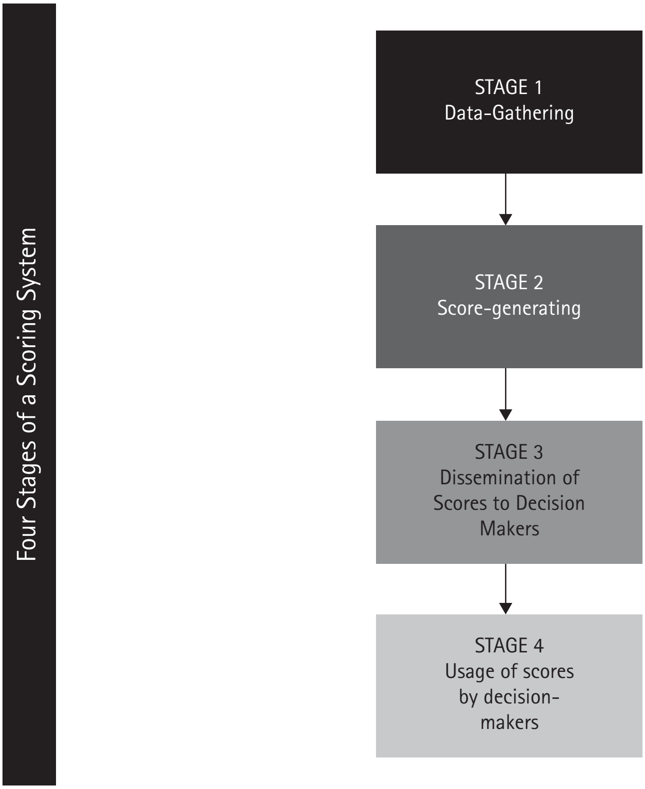
Four Stages of a Scoring System
STAGE 1
Data-Gathering
STAGE 2
Score-generating
STAGE 3
Dissemination of Scores to Decision Makers
STAGE 4
Usage of scores by decision-makers

FIGURE 13.1: Four stages of a scoring system.

Other theorists have offered similar classificatory breakdowns which focus more specifically on the question that interests me in this chapter, that is, the way in which these systems might undermine/promote individual freedom. Gal (2018) argues that there are at least four different kinds of algocratic system, each of which has a distinctive effect on individual choice. The four kinds are (1) *"stated preference"* systems, in which the human users specify exactly what they want the system to do and the system assists in achieving this outcome; (2) *"menu of preferences"* systems, in which the human user doesn't specify their preferred outcome but chooses from a menu of options provided to them by the algorithm; (3) *"predicted preference"* systems, in which the system, based on data-mining (from a large sample population), tries to predict what an individual user will want and target options at them accordingly; and (4) "*self-restraint preference*" systems, in which the algorithm functions as a pre-commitment device, favoring the user's long-term interests (perhaps stated; perhaps predicted) over their immediate interests. As you might imagine, these different kinds of algocratic system have different consequences for individual autonomy. A stated preference algorithm, for example, might seem to be obviously freedom-promoting; a predicted preference algorithm much less so.

In a similar, but more complex, vein, Yeung (2018) tries to develop a taxonomy of algocratic systems. This taxonomy focuses on three main variables that determine the form that an algocratic system can take. Each of these three variables has two "settings," making for eight possible forms of algocracy. The first dimension concerns the nature of the algorithm itself. Is it fixed or adaptive? The second dimension concerns the way in which the algorithmic system monitors individual behavior. Does it "react" to the user's violation of its behavioral standards or does it try to predict and pre-empt the user? The third dimension concerns the role that human regulators play in the system. Does the system automatically enforce its standards (perhaps giving humans a veto power) or does it simply recommend (perhaps strongly) enforcement options to them? Again, the different settings on each of these dimensions would appear to be relevant when it comes to assessing the impact of these systems on individual choice and autonomy. Intuitively, it seems like a system that anticipates and pre-empts violations of prescribed standards, and that automatically enforces sanctions on those violations, poses more of a threat to freedom than a system that simply reacts and recommends. But, again, being sensitive to this complexity is key in any analysis of the freedom-promoting or freedom-undermining effect of algocracy.

Each of these attempts to bring order to complexity has some value to it. Nevertheless, I think there is another way of doing this that is both more illuminating and more revelatory when it comes to evaluating the impact of algocracy on freedom. This method of bringing order to complexity is inspired by the "logical space" method of List and Vallentini (discussed in the previous section) and builds upon the insights provided by all the thinkers mentioned in the previous paragraphs of this section. It starts by identifying three major variables that determine the form that algocratic systems take.

The first is the particular *domain* or type of decision making that is affected by the system. As already mentioned, algocracies arise in different contexts, including financial, governmental, legal, personal, medical and so on. Within each of these contexts many different decisions have to be made, for example, decisions about granting loans, investing in shares, allocating welfare benefits, identifying tax cheats, picking which movie to watch next, deciding when to exercise and how, and so on. The possible variation in affected choices is vast. Indeed, it is so vast that it cannot be easily captured in a formal model or conceptual framework. This is why I essentially ignore it for now. This is not because it is unimportant: when figuring out the freedom-promoting or undermining effects of any particular algocratic decision-making procedure, the domain of decision making should always be specified in advance and the relative importance of that domain should be remembered. This is something I emphasize again later in this chapter. For the time being, however, I set it to one side.

The second variable concerns the main components of the decision-making "loop" that is utilized by these agencies. I mentioned Zarsky, Citron, and Pasquale's attempts to identify the different "stages" in algocratic decision procedures. One thing that strikes me about the stages identified by these authors is how closely they correspond to the stages identified by authors looking at automation and artificial intelligence. For instance, the collection, processing and usage stages identified by Zarsky feel very similar to the sensing, processing and actuating stages identified by AI theorists and information systems engineers. This makes sense. Humans use their intelligence to make decisions and algocratic systems are largely intended to replace or complement human decision makers. It would, consequently, make sense for these systems to break down into those distinct task stages as well. Using the direct analogy with intelligence, I think we can identify four distinct processes undertaken by any algocratic system:

1. *Sensing*: the system collects data from the external world.
2. *Processing*: the system organizes that data into useful chunks or patterns and combines it with action plans or goals.
3. *Acting*: the system implements its action plans.
4. *Learning*: the system uses some mechanism that allows it to learn from what it has done and adjust its earlier stages.

These four processes provide a more precise characterization of the decision-making "loop" that humans can be in, on, or off. The important point in terms of mapping out the logical space of algocracy is that algorithmically coded architectures could be introduced to perform one or all of these four tasks. Thus, there are subtle and important qualitative differences between the different types of algocratic system, depending on how much of the decision-making process is taken over by the computer-coded architecture.

In fact, it is more complicated than that and this is what brings us to the third variable. This one concerns the precise relationship between humans and algorithms for each task in the decision-making loop. As I see it, there are four general relationship-types that could arise: (1) humans could perform the task entirely by themselves; (2)

	(1) Humans perform task	(2) Task is shared with algorithm	(3) Algorithms perform task; Humans supervise	(4) Algorithms perform task; No human input
Sensing	Y or N?	Y or N?	Y or N?	Y or N?
Processing	Y or N?	Y or N?	Y or N?	Y or N?
Acting	Y or N?	Y or N?	Y or N?	Y or N?
Learning	Y or N?	Y or N?	Y or N?	Y or N?

FIGURE 13.2: Sample grid used to classify algocratic systems.

humans could share the task with an algorithm; (3) humans could supervise an algorithmic system; and (4) the task could be fully automated, that is, completely under the control of the algorithm.

Using these second and third variables, we can construct a grid which we can use to classify algocratic systems. The grid looks something like Figure 13.2.

This grid tells us that when constructing or thinking about an algocratic system we should focus on the four different tasks in the typical intelligent decision-making loop and ask of each task: how is this task being distributed between the humans and algorithms? When we do this, we see the "logical space" of possible algocratic systems opening up before us.

Understanding algocracy in this way has a number of virtues. First, it captures some of the true complexity of algocracy in a way that existing conceptual frameworks do not. It not only tells us that there is a large logical space of possible algocratic systems; it allows us to put some numbers on it. Since there are four stages and four possible relationship-types between humans and computers at those four stages, it follows that there are 4^4 possible systems (i.e., 256) *within any given decision-making domain*. That's a minimum level of complexity. You could also make the logical space more complex by adding further dimensions of variance, depending on how fine-grained you want your analysis of algocracy to be. For instance, computer scientists sometimes distinguish between algorithmic processes that are (1) interpretable and (2) non-interpretable (i.e., capable of being deconstructed and understood by humans or not). That could be an additional dimension of variance since at each stage in the decision-making process humans could be sharing a task with an interpretable or non-interpretable system. This would mean that for each stage in the decision-making process there are eight possible configurations, not just four. That would give us a logical space consisting of 8^4 possibilities.

Another virtue of the logical space model is that it gives us an easy tool for coding the different possible types of algocratic system. For the initial two-dimensional model, I suggest that this be done using square brackets and numbers. Within the square brackets there would be four separate number locations. Each location would represent one of the four decision-making tasks. From left-to-right this would read: [sensing; processing; acting; learning]. You then replace the names of those tasks with numbers ranging from 1 to 4 and these numbers could then represent the way in which the task is distributed between the humans and algorithms. A value of "1" would be used when the relevant task is performed entirely by humans, and so on. As follows:

> [1, 1, 1, 1] = Would represent a non-algocratic decision procedure, that is, one in which all the decision-making tasks are performed by humans.
>
> [2, 2, 2, 2] = Would represent an algocratic decision procedure in which each task is shared between humans and algorithms.
>
> [3, 3, 3, 3] = Would represent an algocratic decision procedure in which each task is performed entirely by algorithms, but these algorithms are supervised by humans with some possibility of intervention/veto.
>
> [4, 4, 4, 4] = Would represent an pure algocratic decision procedure in which each task is performed by an algorithm, with no human oversight or intervention.

If we wanted to use a more complicated three-dimensional logical space, we could simply modify the coding system by adding a letter after each number to indicate the additional variance. For example, if we adopted the interpretability/non-interpretability dimension, we could add "i" or "ni" after each number to indicate whether the step in the process was interpretable (i) or not (ni). As follows:

> [4i, 4i, 4i, 4i] = Would represent a pure algocratic procedure that is completely interpretable
>
> [4i, 4ni, 4i, 4ni] = Would represent a pure algocratic procedure that is interpretable at the sensing and acting stages, but not at the processing and learning stages.

This coding mechanism has some practical advantages. Three are worth mentioning. First, it gives designers and creators of algocratic systems a quick tool for figuring out what kind of system they are creating and the potential challenges that might be raised by the construction of that system. Second, it gives researchers something to use when investigating real-world algocratic systems and seeing whether they share further properties (such as their freedom-undermining or promoting potential). For instance, you could start investigating all the [3, 3, 3, 3] systems across various domains of decision making and see whether the human supervision is active or passive across those domains and then trace out the implications of this for individual freedom. Third, it could give us a simple tool for measuring how algocratic a system is or how algocratic it becomes over time. So we might be able to say that a [4ni, 4ni, 4ni, 4ni] is more algocratic than a [4i, 4i, 4i, 4i] and we might be able to spot the drift towards more algocracy within a decision-making domain by recording the changes in the values. This could also be useful when thinking about the freedom-promoting or undermining potential of an algocratic system. As a rough rule of thumb, the more algocratic a system is, the more it is likely to undermine freedom, at least within a given decision-making domain.

This is not to say that there are no problems with the logical space model. The most obvious is that the four stages and four relationships are not discrete in the way that the model presumes. To say that a task is "shared" between a human and an algorithm is to say something imprecise and vague. There may be many different possible ways in which to share a task. Not all of them will be the same. This also is true for the description of the tasks. "Processing," "collecting," and "learning" are all complicated real-world tasks. There are many different ways to process, collect, and learn. That additional complexity is missed by the logical space model. But all conceptual models involve some abstraction and simplification of reality, and all conceptual models miss some element of variation. List and Vallentini's logical space of freedom, for instance, involves a large amount of abstraction and simplification. To say that theories of freedom vary along modal and moral dimensions is to say something vague and imprecise. Specific theories of freedom will vary in how modal they are (i.e., how many possible worlds they demand the absence of interference in) and in their understanding of what counts as a morally legitimate interference. As a result of this, List and Vallentini argue that the logical space of freedom should be viewed as a "definitional schema"—something that is fleshed out in more detail with specific conceptualizations of the four main categories of freedom. The logical space of algocracy can be viewed in a similar light.

Another obvious problem with the logical space model is that it is constructed with an eye to a particular set of normative challenges posed by algocracy. By placing the emphasis on the different ways in which tasks are shared between humans and algorithms, we are naturally drawn to considering the impacts on human agency and autonomy. This means that the model is relatively silent about some of the other normative concerns one could have about algocratic systems (e.g., bad data, biased data, negative consequences). It's not that these concerns are completely shut out or ignored; it's just that they aren't going to be highlighted simply by identifying the location with the logical space that is occupied by any particular algocratic system. What could happen, however, is that empirical investigation of algocratic systems with similar codes could reveal additional shared normative advantages/disadvantages, so that the code becomes shorthand for those other concerns. That said, this limitation of the logical space model is more of a feature than a bug in the present context. This chapter is explicitly focused on the impact of this technology on freedom, and this conceptual framework allows us to do this by giving us a more realistic appreciation of the complexity of algocracy.

4. How Algocracies Can Promote and Undermine Freedom

So freedom is complex and algocracy is complex. It follows that the impact of algocracy on freedom is likely to be complex. When we consider the different dimensions of freedom, and how they might line up with the different possible forms of algocracy, we intuit that there is unlikely to be a simple universal assessment of the impact of the latter on the former. This means we should be suspicious of any arguments that attempt to provide such a general assessment. It also means, unfortunately, that I am not going to be able to provide any definitive analysis of the freedom-undermining or freedom-promoting effects of algocracy in the space of this chapter. Indeed, one of the main conclusions to be reached here is that a definitive analysis is impossible. We need to take each form of algocracy as it comes, looking at how it impacts upon the different dimensions of freedom, and then determining whether this is a good or bad thing, contingent on how we understand the value of freedom. As we do this, we will also need to bear in mind the relative value of freedom across different domains of decision making. It's not necessarily a good thing to have total autonomous control over every decision you make. It may be exhausting or stultifying if you do. So even if we find that some algocratic systems are freedom-undermining in a particular domain, it does not necessarily follow that algocracy is freedom-undermining in general, or that its freedom-undermining effects in that domain are unwelcome.

Despite the difficulties involved, I am going to make some tentative, general, arguments about the possible impact of algocracy on freedom. The intention here is not to offer an unjustified global assessment, but rather to highlight some distinctive challenges, and opportunities, that algocracy might pose for freedom.

Let's consider the challenges first. It should be obvious from the description of how algocratic systems work that they can undermine freedom. If we share or bequeath one or more of the decision-making tasks to an algocratic system, then we open ourselves up to forms of interference and domination that could negatively affect our freedom. We can see this if we take each of the three dimensions of freedom outlined earlier in this chapter (rationality/intelligibility, manipulation and domination) and consider how they may be negatively affected by algocracy.

Recall that the rationality dimension focuses on the extent to which our decision making is the product of conscious and intelligible reflection on our goals and the best way of realizing them through our actions. Algocratic systems obviously threaten the rationality of decision making if they involve complete automation or outsourcing of all decision-making tasks. They also threaten it in more subtle ways, with less pervasive forms of automation, or even when tasks are shared between humans and computers. The non-interpretability (or "epistemic opacity") of algorithmic systems that organize data and make recommendations to humans would undermine rationality to at least some degree. It would mean that we are less certain of the reasons for our actions.

A recommendation is made, but we are not sure why we should follow it. In an extreme form, this can result in humans being "programmed" to act like "simple machines." This is one of the major arguments of Brett Frischmann and Evan Selinger in their book *Re-engineering Humanity* (2018). They give the example of the online contracting environment, as well as the use of app-based services like Google Maps, each of which, they claim, encourages humans to act like simple stimulus-response machines. The algocratic system presents the human user with a stimulus (a box to tick or recommendation to follow) and we give a rote, automatized response. If they are right about this, then even algocratic systems that seem to preserve a degree of human authority may be significantly undermining the rational intelligibility of our decision making. There is no rational reflection on the reasons for our actions; we just blindly follow the instructions.

Closely related to this is the negative impact that algocratic systems can have on the manipulation dimension of freedom. There are many obvious ways in which algocratic systems can manipulate our choices. A system could be designed to coerce you into acting in a certain way. For example, a credit-scoring algocratic system might threaten you with the loss of creditworthiness if you don't act in a prescribed way. There is also the possibility of physical coercion, if the system is joined up with some robotic technology that can physically interfere with the human user. This is not completely far-fetched. The Pavlok behavior change bracelet, for example, is an algocratic system that shocks its user if they don't follow through on certain commitments.[4] For the time being, this system is something that an individual chooses to impose on themselves, not something that is imposed on them by some outside force. Consequently it may not undermine freedom (I return to this in a moment). Nevertheless, it is easy to imagine similar systems being used to physically coerce behavior in a freedom-undermining fashion.

More significant than explicit coercion, however, are the subtle forms of manipulation that are possible through the use of algocratic systems. Yeung (2017) argues that algocratic systems enable "hypernudging," which is a kind of behavior change technique that operates beneath the radar of conscious awareness and happens in a dynamic and highly personalized fashion. Nudging is a concept that was made popular by Cass Sunstein and Richard Thaler (2009). It involves using insights from behavioral science to construct choice architectures that "nudge" people towards actions that are welfare-maximizing or for the common good. For example, setting the default on retirement savings to "opt-out" rather than "opt-in," or placing healthy foods at eye level and unhealthy ones below or above, makes it more likely that people will choose options that are in their long-term interests. Nudges usually operate on subconscious biases in human reasoning. Sunstein and Thaler maintain that nudging is not freedom-undermining because it is still possible for people to identify and reject the "nudges." Others are more doubtful and argue that nudges are highly manipulative (Sunstein 2016). Whatever the merits of nudging, Yeung's point is that algocratic technologies bring nudging to an extreme. Instead of creating a one-size-fits-all choice architecture that is updated slowly, if ever, you can create a highly personalized choice architecture that learns and adapts to an individual user. This can make it much more difficult to identify and reject the nudges.

Finally, there is the potential impact on the domination dimension. Recall that domination arises whenever decision making is subject to the arbitrary will of another. This "other" may not directly manipulate or interfere with your behavior, but the mere fact that they could (in some possible world), and that you have to keep on their good side to avoid any such interference, is enough to compromise your freedom. Hoye and Monaghan (2018) and Graf (2017) both argue that the mass surveillance on which algocratic systems are built enables domination on a mass scale. If your behavior is being monitored and mined for patterns and predictions, then it is possible that some of that behavior might trigger interference from the system itself (thanks to automation) or from some human controller of the system, particularly if it falls outside the normal or permissible range of the system's expectations. This means that you have to live within the constraints established by the system if you want to avoid interference. If we are constantly flitting from the grasp of one algocratic system to the next—across the different domains of life—the extent of freedom-undermining domination could be quite dramatic. It might give rise to what I call "algocratic micro-domination."

"Micro-domination" is a concept that I take from the work of Tom O'Shea (2018), who uses it to understand the forms of domination experienced by people with disabilities. He argues that people with disabilities often suffer from many small-scale instances of domination. If they live in an institutional setting, or are heavily reliant on care and assistance from others, then large swathes of their daily lives may be dependent on the good will of others. They may need these others to help them when they wake up, when they go to the bathroom, when they eat, when they go outside, and so on. Taken individually, these cases may not seem all that serious, but aggregated together they take on a different guise:

> The result is often a phenomenon I shall call 'micro-domination': the capacity for decisions to be arbitrarily imposed on someone, which, individually, are too minor to be contested in a court or a tribunal, but which cumulatively have a major impact on their life.
>
> (O'Shea 2018, 136)

The pervasiveness of algocracy in modern society can give rise to a similar phenomenon. Many small-scale, arguably trivial, choices in our everyday lives take place within algocratic systems: what route to drive, what news stories to read, who to talk to on social media, what film to watch next and so on. A network of devices monitors and tracks our behavior and sends us prompts and reminders. This means that we are now the "subjects" of many algorithmic masters. They surveil our lives and create a space of permissible/acceptable behavior. Everything is fine if we stay within this space. We can live happy and productive lives (perhaps happier and more productive than our predecessors), and to all intents and purposes, these lives may appear to be free. But if we step out of line we may be quick to realize the presence of the algocratic masters. Consider, Janet Vertesi's experiences in trying to "hide" her pregnancy from the algocratic systems that monitor consumer behavior online (Vertesi 2014). Vertesi, an expert in Big Data, knew that online marketers and advertisers like to know if women are pregnant. Writing in 2014, she noted that an average person's marketing data is worth about 10 cents whereas a pregnant person's data is worth about $1.50. She decided to conduct an experiment in which she would hide her own pregnancy from the online data miners. This turned out to be exceptionally difficult. She had to avoid all credit card transactions for pregnancy-related shopping. She had to implore her family and friends to avoid mentioning or announcing her pregnancy on social media. When her uncle breached this request by sending her a private message on Facebook, she deleted his messages and unfriended him (she spoke to him in private to explain that even these private messages are mined for data). In the end, her attempt to avoid algocratic domination led to her behavior being flagged as potentially criminal:

> For months I had joked to my family that I was probably on a watch list for my excessive use of Tor and cash withdrawals. But then my husband headed to our local corner store to buy enough gift cards to afford a stroller listed on Amazon. There, a warning sign behind the cashier informed him that the store "reserves the right to limit the daily amount of prepaid card purchases and has an obligation to report excessive transactions to the authorities."
>
> It was no joke that taken together, the things I had to do to evade marketing detection looked suspiciously like illicit activities. All I was trying to do was to fight for the right for a transaction to be just a transaction, not an excuse for a thousand little trackers to follow me around. But avoiding the big-data dragnet meant that I not only looked like a rude family member or an inconsiderate friend, but I also looked like a bad citizen.
>
> (Vertesi 2014)

Vertesi wouldn't have had any problems if she had lived her life within the space of permissible activity created by the system of algorithmically controlled commerce. She wouldn't have been interfered with or overtly sanctioned. By stepping outside that space, she opened herself up to interference. She was no longer tolerated by the system. This is a good illustration of how algocratic micro-domination might arise.

But it is not all doom and gloom. If designed and implemented in the right way, algocratic systems can promote, rather than undermine freedom. We see this most clearly if we remember that (1) sometimes you may have to tradeoff one dimension of freedom against another and (2) sacrificing freedom in one choice domain may benefit freedom in another. There are consequently two mechanisms, in particular, that algocratic systems could use to promote freedom.

The first is *choice filtration*. In order to make a rationally intelligible decision, you must be able to identify and select among options that might (or might not) be conducive to your goals. It's often assumed in mainstream economic theory that the more options the better (the more likely it is that someone can find an option that satisfies their preferences). But there are some experimental studies in psychology that cast this into doubt. Barry Schwartz, and his colleagues, famously identified the "paradox of choice," which states that if people are confronted with too many options they can become overwhelmed and unable to decide what to do (Schwartz 2004). At a certain extreme, too many options actually undermines freedom. Like many findings in experimental psychology (Open Science Collaboration 2015), this one is under attack for failing to replicate, but the most comprehensive meta-analysis of the phenomenon (Scheibehenne et al 2010) suggests that although it may not exist in every choice context, it does exist in some and often with quite a large effect size. One of the advantages of algocratic systems is that they can help to filter choices and reduce the feeling of being overwhelmed. Certainly, I feel grateful when Netflix recommends viewing options to me. It makes it much easier to use my rationality to select something that is conducive to my goals. More generally, algocratic systems can make decision making more rationally intelligible by bringing order to the chaos of data. By identifying salient patterns and bringing them to our attention, they can give us access to decision-relevant information that we might otherwise lack. This is not true for every algocratic system. Some can result in more opacity, but we must remember that the world is always somewhat opaque to human reason. We don't yet have a theory of everything. Until we do, we must compromise and accept some element of decisional opacity. By illuminating and helping us to make sense of some data, algocratic systems might represent a good compromise when it comes to a minimum level of opacity. This may mean, however, that we have to accept some domination or potential manipulation into our lives—because we have to let the systems monitor us to do their work—but this might be a worthwhile tradeoff.

The second way that algocratic systems can help is by creating *cognitive slack*. This is similar to the first mechanism but it has more to do with increasing overall freedom by offloading some decision-making domains to algocratic systems. The work of Sendhil Mullainathan and Eldar Shafir argues that people who suffer from various kinds of scarcity (e.g., scarcities of time or income) often suffer from impaired cognitive control over their behavior (Mullainathan and Shaffir 2012 and 2014; Shah, Mullainathan and Shaffir 2012). To be more precise, it argues that scarcity places a tax on cognitive bandwidth. "Bandwidth" is general term they use to describe the ability to focus on tasks, solve problems, exercise control, pay attention, remember, plan, and so on. All of these things are, of course, crucial to freedom. Mullainathan and Shafir's main contention, backed up by a series of experimental and field studies, is that scarcity narrows cognitive bandwidth. For example, if you have less money, you tend to be uniquely sensitive to stimuli relating to price. This leads to a cognitive tunneling effect: you become very good at paying attention to anything relating to money in your environment, but have reduced sensitivity to everything else. This results in less fluid intelligence and less executive control, which means you are less able to make rationally intelligible autonomous decisions across the full range of life's activities. Miles Brundage and I (Brundage and Danaher 2017) have argued that algocratic systems could address this problem, at least in part. By offloading decision-making authority to an algocratic system you can free up some "cognitive slack," which means you can escape from your cognitive tunnel, and promote freedom in other aspects of your life. This is, admittedly, just a hypothesis, but it is one worth exploring in more detail.

So we have then five general mechanisms through which algocracy might affect freedom, three of them negative and two positive. There are probably many more mechanisms that have yet to be created, identified, and debated. Let me just close by injecting one further note of caution into the discussion. In addition to remembering the complex relationship between freedom and algocracy, it is also worth remembering, and avoiding, the "status quo" bias in how we think about that relationship (Bostrom and Ord 2006). There is always a danger in thinking about novel technologies that we fixate on their "newness." We look at the "new" threats or opportunities they pose for cherished values and assume that they must be either terrible or wonderful. We ignore the fact that new technologies do not arise in a vacuum. They emerge into a world that already has its own baseline mix of threats and opportunities. The impact of the new technology must always be evaluated relative to that pre-existing baseline. So when it comes to freedom and algocracy, we need to remember that things like manipulation and domination are nothing new. The world is already replete with them. We need to figure out whether algocratic technologies make things better or worse.

5. Conclusion

To sum up, in this chapter I've presented three frameworks for thinking about freedom in an age of algocracy. First, I've outlined a framework for thinking about the value of freedom. I've argued that freedom is a complex value. Indeed, it is so complex that it is best to think of it as a scalar and multi-dimensional value, something that you can have more or less of, rather than something you either have or you don't. It is something about which you may need to compromise, by trading the different dimensions of freedom off against one another, and not trying to maximize across all dimensions at once. I've also argued that there are several ways in which to value freedom (intrinsically, instrumentally etc.) and that these always need to be factored in when thinking about the impact of algocratic systems on our freedom.

Second, I've outlined a framework for thinking about the nature of algocracy. Again, I've argued that algocracy is a complex phenomenon. Algocratic systems arises in many different domains and within any given domain there is a large, logical space of possible forms that the system could take. These forms vary depending on how they share and distribute decision-making tasks between humans and machines. It is important to remember these different possible forms of algocracy, both from a design perspective and from a critical perspective, because some pose a greater threat to freedom than others.

Finally, I've outlined a framework for thinking about the likely impact of algocracy on freedom. Given the complexity of freedom and the complexity of algocracy, I've argued that there is unlikely to be a simple global assessment of the freedom-promoting or undermining power of algocracy. This is something that has to be assessed and determined on a case-by-case basis. Nevertheless, there are at least five interesting and relatively novel mechanisms through which algocratic systems can both promote and undermine freedom. We should pay attention to these different mechanisms, but do so in a properly contextualized manner, and not by ignoring the pre-existing mechanisms through which freedom is undermined and promoted.

Notes

1. Brief aside: one could argue that I should use the term "complicated" as opposed to "complex." The latter has a particular meaning in the field of complexity science. It is used to describe the fundamental unpredictability of certain systems containing many distinct interacting parts. The former describes a system that consists of many parts but is relatively easy to predict and manage. Although I am not concerned here with the predictability or explainability of a phenomenon, it could be argued that when I say that freedom is "complex" what I really mean is that it is "complicated," i.e., consists of many parts.
2. Pettit would probably dispute the way in which List and Vallentini characterize his theory. His theory holds that only arbitrary forms of domination are freedom undermining, but he tries to define arbitrariness in non-moral terms. List and Vallentini argue that it is very difficult for him to do this and discuss this issue at length in their paper.
3. There are some similarities here with Lawrence Lessig's idea, drawn from legal theory, that "code is law," but there are differences too. Lessig's "code is law" idea has a narrow application (Lessig 1999; 2006).
4. See https://pavlok.com

References

Aneesh, A. 2006. *Virtual Migration*. Durham, NC: Duke University Press.

Aneesh, A. 2009. "Global Labor: Algocratic Modes of Organization." *Sociological Theory*, 27, no. 4: 347–370.

Binns, Reuben. 2018. "Fairness in Machine Learning: Lessons from Political Philosophy." *Journal of Machine Learning Research* 81: 1–11.

Bostrom, Nick, and Ord, Toby. 2006. "The Reversal Test: Eliminating Status Quo Bias in Applied Ethics." *Ethics* 116, no. 4: 656–679.

Brundage, Miles, and Danaher, John. 2017. "Cognitive Scarcity and Artificial Intelligence: How Intelligent Assistants Could Alleviate Inequality." *Philosophical Disquisitions*, May 15. http://philosophicaldisquisitions.blogspot.com/2017/05/cognitive-scarcity-and-artificial.html?

Burrell, Jenna. 2016. "How the Machine Thinks: Understanding Opacity in Machine Learning Systems." *Big Data and Society* (January–June): 1–12. doi:10.1177/2053951715622512

Carter, Ian. 1995. "The Independent Value of Freedom." *Ethics* 105, no. 4: 819–845.

Carter, Ian. 1999. *A Measure of Freedom*. Oxford: OUP.

Citron, Danielle & Pasquale, Frank. 2014. "The Scored Society: Due Process for Automated Predictions." *Washington Law Review*, 89, no. 1: 1–34.

Crawford, Kate. 2013. "The Hidden Biases of Big Data." *Harvard Business Review*, April 1. https://hbr.org/2013/04/the-hidden-biases-in-big-data.

Danaher, John. 2015. "How Might Algorithms Rule Our Lives? Mapping the Logical Space of Algocracy." *Philosophical Disquisitions* June 15. https://philosophicaldisquisitions.blogspot.com/2015/06/how-might-algorithms-rule-our-lives.html

Danaher, John. 2016a. "The Logical Space of Algocracy: A Guide to the Territory." Talk to the IP/IT/Media Law Discussion Group, Edinburgh University School of Law, November 25.

Danaher, John. 2016b. "The Threat of Algocracy: Reality, Resistance and Accommodation." *Philosophy and Technology*, 29, no. 3: 245–268.

Danaher, John. 2018. "Moral Freedom and Moral Enhancement: A Critique of the 'Little Alex' Problem." *Moral Enhancement: Critical Perspectives*, edited by Michael Hauskeller and Lewis Coyne. Cambridge: Cambridge University Press.

Dworkin, Gerald. 1988. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Eubanks, Virginia. 2018. *Automating Inequality*. New York: St Martin's Publishing.

Ferguson, Andrew. 2017. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law*. New York: New York University Press.

Frischmann, Brett, and Selinger, Evan. 2018. *Re-Engineering Humanity*. Cambridge: Cambridge University Press.

Gal, Michal. 2018. "Algorithmic Challenges to Autonomous Choice." *Michigan Journal of Law and Technology* 25 (Fall): 59–104.

Gräf, Eike. 2017. When Automated Profiling Threatens Freedom: a Neo-Republican Account. *European Data Protection Law Journal* 4: 1–11.

Hoye, J. Matthew, and Monaghan, Jeffrey. 2018. "Surveillance, Freedom and the Republic." *European Journal of Political Theory* 17, no. 3: 343–363.

Killmister, Jennifer. 2017. *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance*. London: Routledge.

Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. New York: Basic Books.

Lessig, Lawrence. 2006. *Code 2.0*. New York: Basic Books.

List, Christian. and Vallentini, Laura. 2016. "Freedom as Independence." *Ethics* 126, no. 4: 1043–1074.

Mullainathan, Sendhil, and Shafir, Eldar. 2012. *Scarcity: The True Cost of Not Having Enough*. London: Penguin.

Mullainathan, Sendhil, and Shafir, Eldar. 2014. "Freeing Up Intelligence." *Scientific American Mind* 25, no. 1 (Jan/Feb): 58–63.

Noble, Safiya. 2018. *Algorithms of Oppression*. New York: New York University Press.

Open Science Collaboration. 2015. "Estimating the Reproducibility in Psychological Science." *Science* 349, no. 6251: aac4716.

O'Neil, Cathy. 2016. *Weapons of Math Destruction*. London: Penguin.

O'Shea, Tom, 2018. "Disability and Domination." *Journal of Applied Philosophy* 35, no. 1: 133–148.

Pasquale, Frank. 2015. *The Black Box Society*. Cambridge, MA: Harvard University Press.

Pettit, Philip. 2001. *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press.

Pettit, Philip. 2011. "The Instability of Freedom as Non-Interference: The Case of Isaiah Berlin." *Ethics* 121, no. 4: 693–716.

Pettit, Philip. 2014. *Just Freedom: A Moral Compass for a Complex World*. New York: W.W. Norton and Co.

Polonetsky, Jules, and Tene, Omar. 2013. "Privacy and Big Data: Making Ends Meet." *Stanford Law Review* 66: 25–33.

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.

Rouvroy, Antoinette. 2013. "Algorithmic Governmentality and the End (s) of Critique." *Society of the Query*, 2. https://networkcultures.org/query/2013/11/13/algorithmic-governmentality-and-the-ends-of-critique-antoinette-rouvroy/

Rouvroy, Antoinette. 2015. "Algorithmic Governmentality: a Passion for the Real and the Exhaustion of the Virtual." *All Watched over by Algorithms*. Berlin. Conference Presentation, 29th January. https://www.academia.edu/10481275/Algorithmic_governmentality_a_passion_for_the_real_and_the_exhaustion_of_the_virtual

Scheibehenne, B., Greifeneder, R., & Todd, P. M. 2010. "Can There Ever Be Too Many Options? A Meta-analytic Review of Choice Overload." *Journal of Consumer Research*, 37: 409–425.

Schwartz, Barry. 2004. *The Paradox of Choice: Why Less Is More*. New York, NY: Harper Collins.

Shah, Anuj. K., Mullainathan, Sendhil, & Shafir, Eldar. 2012. "Some Consequences of Having Too Little." *Science*, 338, no. 6107: 682–685.

Skinner, Quentin. 2008a. *Hobbes and Republican Liberty*. Cambridge: Cambridge University Press.

Skinner, Quentin. 2008b."The Genealogy of Liberty." Public Lecture, UC Berkley, 15 September.

Skinner, Quentin. 2012. *Liberty before Liberalism*. Cambridge: Cambridge University Press.

Sunstein, Cass. 2016. *The Ethics of Influence*. Cambridge, UK: Cambridge University Press.

Thaler, Richard, & Sunstein, Cass. 2009. *Nudge: Improving Decisions about Health, Wealth and Happiness*. London: Penguin.

Vertesi, Janet. 2014. "Internet Privacy and What Happens When You Try to Opt Out." *Time*, May 1.

Yeung, Karen. 2017. " 'Hypernudge': Big Data as a Mode of Regulation by Design." *Information, Communication and Society*, 20, no. 1: 118–136.

Yeung, Karen. 2018. "Algorithmic Regulation: A Critical Interrogation." *Regulation and Governance* 12, no. 3: 505–523.

Zarsky, Tal. 2012. "Automated Predictions: Perception, Law and Policy." *Communications of the ACM* 15, no. 9: 33–35.

Zarsky, Tal. 2013. "Transparent prediction." *University of Illinois Law Review* 4: 1504.

CHAPTER 14

(BIO)TECHNOLOGY, IDENTITY, AND THE OTHER

ANNA GOTLIB

1. Introduction

Technological innovation, whatever its forms, has occupied both the imagination and the efforts of generations, whether in the pursuit of quotidian human convenience, or in the more abstract goals of pushing the edge of what is conceivable. While the dawn of the twenty-first century has brought nothing if not an acceleration of the pace and complexity of these ever-evolving technologies, it has also widened the schism between what I call techno-optimism and techno-pessimism—the views that these innovations will in significant ways either save humanity (from disease, from external threats, from itself), or else destroy it if not in body, then definitely in many of the ways that matter. This paper makes a case for a third position—techno-incrementalism—that argues for caution, and for an optimism tempered by care about empirical facts, consequences, and moral worries about the ongoing re-making of our selves. It focuses on several specific (bio)technologies currently at the forefront of both scientific and philosophical discourses, and offers a critique of the often unearned and uncritical reception that they receive from professionals and the lay public alike. I say "uncritical" not because I take techno-pessimism to be the proper attitude toward innovation—indeed, I take technological innovation to be necessary to human development. However, I do believe that unbridled techno-optimism can also place us at risk for innovating beyond our ability to fruitfully incorporate our new inventions into a well-lived human life, to contribute to the sense of human well-being. That is, we ought to look, hard and carefully, before we leap.

Yet the desire for biomedical enhancement is a powerful one. Although a number of critics have focused on our competitive, striving selves that demand technological advancement (Smith 2014; Kenwright 2018), I suggest a reason that is perhaps a bit more modest. Indeed, while the stories that human beings might tell each other, and themselves, about why they want to be different—to be "better"—may differ, a common thought underwrites many of them: the wish to be more at home in the world and in one's body; to be more comfortable and capable navigating the numerous challenges with which life presents us; to be able to cope with whatever befalls us that we could not predict; to simply *suffer less*. And so we seek novel technological solutions for the individual and communal struggles inherent in finding ourselves in an increasingly complex, demanding, and often unforgiving world that thrives on competition, hierarchies, and expediency while making little allowance for the idiosyncrasies, abilities, and needs of diverse individuals and groups. Least of all is the modern human allowed the time and space for self-reflection, considerations of larger meanings, (inter)dependencies, for questions of personal identity itself. But these things matter enormously. This paper addresses some of the reasons why.

Specifically, we ought to consider the impact of the nature, function, and role of evolving technology on personal and group identities, and especially the identities of vulnerable others. Difficult questions remain about how these technologies, while in many ways positive, can also be sources of further othering of individuals and populations, and about how the promise of human progress through enhancement can also lead to a deepening human liminality. I begin by considering three unique areas of biotechnological innovation: First, I discuss developments in neuroscience, primarily research addressing memory modification. Second, I examine the recent innovations in virtual reality environments. Third, I focus on robotic assistance for the elderly. Finally, I suggest some ways in which these technologies could be re-envisioned as less othering, more inclusive, and less damaging to our sense of who we are. While my arguments are not strictly directed against the further growth of these technologies—in fact, I think that they have much to offer that is an overall social good—they are precautionary in nature. But the direction in which we ought to be heading as a technological species, I suggest, should be motivated as much by considerations of powerful othering and identity-constituting effects of technological innovations as by the adventurism of techno-evolution itself.

2. Manipulating Memories, Othering the Self

2.1 The Quest to Forget

Our memories define us. Our identities, our selves, our lives would be unrecognizable without them. But when things go very wrong—in times of war, or after acts of violence, and other physical and psychological wounds—these same memories can become not just a burden, but a debilitating barrier, an ongoing trauma that haunts the sufferer daily. The result is self-estrangement, as well as estrangement from others—a sense of irreparability of the self, the inescapability of one's demons. In the face of such ongoing trauma, cognitive and behavioral therapy are not only too often ineffective, but lead to devastating relapses, often leaving the patient broken, rudderless, and hopeless about any possible relief (Bell 2008). It is then that we begin asking whether there is a way to avoid the suffering, to lessen the trauma—in other words, to *forget*.

These memories of fear, love, hatred, despair—the powerful ones that tend to dominate our self-concept—differ from implicit, procedural memories (the "how to" memories of how to drive or how to tie one's shoe, for example); or from the conscious, visual memories of, say, appointments, or the words to a favorite song, which can be found in the hippocampus. The manipulation of these more powerful memories has been the response to these memory-related fears and anxieties of the biomedical and neuroscientific communities (Miller 2004; Glannon 2006; Donovan 2010; Gräff et al. 2014).

Memory manipulation is a general term for a number of developments in neurobiological research (Evers 2007). Very generally, there are two kinds of memory manipulation technologies that I address here: (1) neurobiological memory blunting, and (2) memory erasing, currently in various trials. I suggest that they both are grounded in the idea that memory-driven, identity-constituting narratives, whether of individuals or groups, are editable, revisable, and, importantly, now open to a kind of deliberate fictionalization not available to us at any previous point in human history. I emphasize the "deliberate" in part because fictionalization of both our memories and our narratives is quite common: indeed, we forget, confuse, lie, change ourselves into the wronged protagonist rather than the offending villain, and so on, all the time. However, it is one thing to forget—because forgetting is what human beings do all the time—and quite another to deliberately blunt or remove unwanted memories in a way that is similar to the moral difference between deliberately lying (while still knowing the truth) and remembering incorrectly. I thus suggest that it makes a moral and political difference when these practices of memory manipulation are granted the normative imprimatur of powerful biomedical institutions. And as current research has taken us from a place of wondering about the *Eternal Sunshine of the Spotless Mind* (Gondry et al. 2004) to a place of its reification, the phenomenon of memory modification is precisely the sort of bio-adventurism that calls attention to itself as a paradigm-changer, as something that alters the conditions of possibility of how we conceive of the world and of ourselves.

Before proceeding any further, a few disclaimers: My note of caution about memory manipulation is not in any way intended to ignore, or set aside, the very real suffering of those with PTSD and related conditions, nor is it to suggest that research that tampers with the contents of, and access to, our memories needs be stopped. My argument is simply that we proceed carefully, incrementally, and, most importantly, that we pause just enough to ask the necessary questions about the relationship between memory tampering and personal identity and other worries about where the biomedical and the moral tend to intersect. And although I will set aside questions about the importance of collective memory—this complicated issue deserves its own separate treatment and is a bit outside of the scope of this paper—I will focus on worries about how memory manipulation might affect our concept of who we are, and thus of the moral and political significance of our personal identities.

So, what specifically is meant by memory manipulation? Primarily, it is about two things; attenuating the relationship between our emotions and our memories (thus avoiding the traumatizing effects of conditions like posttraumatic stress disorder; PTSD), and, more radically, removing troubling memories altogether. The idea of a manufactured lacunar amnesia (the loss of memory about one specific event) unites both of these possibilities, as it leaves a gap in one's memory.

> Some scientists now believe that memories effectively get rewritten every time they're activated. Studies on rats suggest that if you block a crucial chemical process during the execution of a learned behavior—pushing a lever to get food, for instance—the learned behavior disappears. The rat stops remembering. Theoretically, if you could block that chemical reaction in a human brain while triggering a specific memory, you could make a targeted erasure. Think of a dreadful fight with your girlfriend while blocking that chemical reaction, and zap! The memory's gone (Chadwick 2004).

Thus, two unique processes of memory modification are currently being investigated: First, research has demonstrated that the off-label use of the beta-blocker propranolol, if taken a short time after a traumatic event, can reduce the intensity of a given memory and the subsequent risk of PTSD. It attenuates one's emotional reaction to unwanted, difficult memories in ways that make living with those memories more bearable without erasing them altogether. When we recall a long-term memory, the memory is un-written and re-written—this process is reconsolidation. Propranolol administered during reconsolidation attenuated the connection between the memory and the emotional reaction to the memory (Parens 2010; Chandler et al. 2013).

Second, researchers in Brooklyn's SUNY Downstate Medical Center (and at other research facilities) have recently delivered a single dose of an experimental drug to that part of the brain which deals with specific memories, emotions, motor skills, and spatial relations, blocking the activity of a substance that is required for the brain to retain what it has learned. And although the experiment has only been attempted using animals, this research might very well work on people. (Carey 2009). More recently, new approaches to total bad memory blocking are being developed in ways that invoke the plasticity of the brain—its ability to forget a fear reaction, and thus a fearful memory, in response to external rearranging of a given set of neural connections:

> [In] a 2014 study . . . a team from RIKEN-MIT Center for Neural Circuit Genetics were able to change bad memories to good in male mice. Using a technique called optogenetics, in which genetically encoded, light-responsive proteins are inserted into cells, the scientists were able to pinpoint where a mouse's negative memory of a shock to the foot was formed, in the neural circuitry that connects the dentate gyrus in the hippocampus to the amygdala. The researchers then manipulated those

> neurons with lasers. Each time the mice ventured to a certain part of their enclosure, the negative memory was reactivated, and they quickly learned to fear the area.
>
> The male mice were then allowed to frolic with female mice, while the same neurons were tapped, helping to switch their messages from one of pain to one of pleasure. The next time the male mice ventured into the chamber, their fear reactions had vanished (Lu 2015).

In an earlier study that worked to identify the neurons supporting a particular memory, "the resulting memory loss was robust and persistent, which suggests that the memory was permanently erased. These results establish a causal link between a specific neuronal subpopulation and memory expression, thereby identifying critical neurons within the memory trace" (Han, Kushner, Yiu et al. 2009).

Quite certainly, this is biomedical and neuroscientific progress which offers some hope for those caught in the trap of PTSD and related traumatic conditions. What is less certain is whether we have taken the time and care to examine some of the more problematic aspects of memory manipulation. I now turn to a brief consideration of this worry.

2.2 Deliberate Forgetting as Othering

We often want to control our memories—sometimes badly enough to change them. Even before biomedical science has revealed possibilities for doing so, we have been fantasizing how it might take place. In the film *Eternal Sunshine of the Spotless Mind*, Joel Barish is treated by Dr. Mierzwiak, the head of a local memory erasure clinic, Lacuna Inc., in his attempt to erase not only a bad break-up, but also the entire existence of his ex-lover, Clementine Kruczynski, after finding out that she had done the same to him. As the procedure progresses, Joel begins to regret his decision, but by that time, it is too late, and Clementine fades from his memory.

In Paul Verhoeven's (Schwarzenegger et al. 2001) film *Total Recall* (a loose adaptation of Philip K. Dick's short story, "We Can Remember It for You Wholesale"), it is the year 2084, and Douglas Quaid, a seemingly happily married man, is beset by recurring dreams about Mars, where he has never been, and a woman he has never seen. Deciding that he needs a "vacation" from his life as a construction worker, he visits Rekall, a company which specializes in selling imaginary vacations by implanting false memories, and purchases a fantasy vacation on Mars, where he will instead be a spy. However, the Rekall procedure seems to unlock some part of his psyche that belies his notions of himself as an ordinary family man, and reveals that, indeed, he is not who he thinks he is at all. In fact, Quaid comes to find out that the past eight years of what he has taken to be his "real life" are nothing but more inserted false memories, and he himself is actually a secret double agent named Hauser. In this way, Douglas Quaid's biographical identity and his sense of personal meaning are not just broken—he literally does not know who, what, or why he is.

In both cases, the intuition that I am after here is this: although we are certainly bothered by voluntary memory modification undertaken for what some might consider frivolous reasons, such as a bad love affair, and while we are bothered by Rekall-style third-party machinations for nefarious purposes, we might be most bothered by the freedom, *the choice*, to alter one's memory altogether. I want to offer two reasons to be a bit suspicious of memory manipulation: first, a potential for othering the self; second, a potential for othering of others. Before turning to these two claims, however, a brief summary of the memory modification debate thus far.

Ever since propranolol began to gain some popularity in treating PTSD and other memory-implicating disorders, theorists from a number of disciplines have disagreed about its technical and moral merits. Some, like the President's Council on Bioethics, have viewed this new technology as ethically worrisome—something akin to Homer's lotus flowers, which made Odysseus's men forget everything they ever knew and, indeed, made them somewhat less human (President's Council on Bioethics and Kass 2003). Others worried about the overmedicalization of trauma—of medicalizing memories themselves (Hurley 2007). Still others took the opposing view and suggested that "as long as individuals using MMTs [memory modifying technologies] do not harm others and themselves [. . .]and as long as there is no prima facie duty to retain particular memories, it is up to individuals to determine the permissibility of particular uses of MMTs." (Liao and Sandberg 2008, 96). Finally, arguments were voiced, claiming that the approach to "general" trauma, and traumatic memories, did not take sufficient account of the idiosyncrasies of particular circumstances (such as, for example, memories of sexual assault versus memories of war). Any criticism or support of memory modifying technologies, these arguments noted, needs to be tested in specific cases before we are able to have any confidence in our conclusions about their moral merit (Chandler, Mogyoros, Rubio, and Racine 2013).

So who is right? I now return to my two objections to memory modification via deliberate forgetting: the othering of oneself and the othering of others. My concerns about the othering of the self echo those of David Wasserman, who focused on the break with our past—and, in fundamental ways, with our identities—that such technologies might make possible (Wasserman 2004). But how—and why?

The answer has something to do with how our selves are constituted by examining the relation of memory to identity. If we assume for a moment a narrative view of personal identity (Lindemann 2001; Charon et al. 2002; Gotlib 2014), then first, second, and third-personal stories told by and about us, create this idea of who we are, what we can do (and why), to what we are entitled, what is expected of us, and so on. As Marya Schechtman notes, this narrative identity includes "a sequential listing of life events [and] a story of how the events in one's history lead to other events in that history" (Schechtman 2007, 160). In other words, one's life is not an episodic one—it is not constituted by atomistic, unrelated events—but is diachronic, requiring narrative explanations of how its events, thoughts, actions, and so on are connected to each other, and to a coherent version of a self (Gotlib 2014). For this, one requires memories. Moreover, "having a narrative is necessary for engaging in certain sorts of complex, person-specific activities . . . for instance autonomy, moral agency, prudential reasoning, or other kinds of higher-order capacities" (Schechtman 2007, 159). For this, too, we need access not only to our roles in our life's events, but a way to connect event-specific knowledge to the rest of our lives in a way that creates not only sense, but meaning—thus memories, again.

Given this sketch of a narratively-conceived notion of personal identity, Wasserman's worry about what I call the othering of the self becomes clear. When we mute, or erase completely, our memories, we might very well remove the pain of PTSD and other traumas. But what we also might do is lose the diachronic connection with ourselves, with the narratives of our own lives. He asks:

> Will dulling the emotions associated with memory encourage a cavalier attitude towards our own pasts? Would anesthetizing ourselves to the recall of such memories make us more or less whole?
>
> (Wasserman 2004, 18)

But the worry does not end with the loss of our emotional connection to our past selves. In choosing to fictionalize who we are by removing the unwanted or feared parts of our life's narrative, we might very well create one with more narrative appeal, one which we endorse because it offers the version of ourselves that we prefer over the one that actually exists. It might be a self that never was assaulted, never fought in a war, never experienced a painful break-up, never became seriously ill. It very well might be a self which sounds like a character we admire—not Lady Macbeth, unable to wash away her crimes, not someone who found herself choosing a less unbearable choice out of untenable possibilities, but someone who did the right thing, who triumphed. Or else someone who did not have to face such choices, and their consequences, at all. Wasserman notes that

> [t]he availability of memory-altering drugs will, so to speak, shift the balance of power in our extended temporal relations to ourselves ever more to the present and future. It will allow us to break or weaken the grip of the past, while leaving us in the present with the uneasy awareness that our capacity to shape ourselves by what we chose to do, experience, and remember is tentative and indefinitely revisable. We will have more freedom to re-invent ourselves, but less security about who we are, or will end up being
>
> (Wasserman 2004, 18).

In other words, while we might gain the freedom to narrate alternative, non-traumatized versions of ourselves into reality, we might also create epistemically spectral selves that bear very little resemblance to either our actual phenomenologies, or to versions of ourselves known to others. Unlike the more quotidian ways in which we, deliberately as well as without intending to, forget, reinterpret, and revise our narratives about who we are, this medicalized forgetting is less about creating a more appealing version of ourselves, and more about constructing a self that is less burdened by all that it does not wish to recall. Yet regardless of the success of these more quotidian narrative manipulations, it seems plausible that unless we are incredibly skilled at self-deception, some part of us will still be the bearer of a more honest story of what took place, despite our own inner protestations. But medically-mediated forgetting seems different: it not only allows the kind of self-fictionalizing that worries Wasserman, but it does so without leaving the kinds of traces of the deliberately-forgotten versions of ourselves that, without such interventions, might still remind us about those unwanted, yet, vital, selves. And so we will remain on shifting sands, with neither others, nor ourselves, knowing whether this new identity before us has anything like a diachronic connection to the past, whether it can tell a coherent story of who it is—or whether it is an other, a selected character sketch, awaiting yet further revision.

As for the othering of others, Elisa Hurley, correctly, worries about

> cut[ting] off access to the emotions experienced at the time of trauma, access that might be important for holding oneself and others accountable for moral wrongdoing (Hurley 2007, 36).

While the clear cases of not desiring criminals to be able to forget their wrongdoing are more obvious, what is less immediately evident, but just as important, is that we do not want to lose the kind of moral language, a moral vocabulary, among ourselves, that allows us to regret bad things that we, and others, have done. Or even perhaps to be bothered, to feel guilt—to remain moved—by failed relationships, mistakes, missteps, and so on. We need bad memories, in other words, not just to stay connected to ourselves, but to each other—empathetically, compassionately, fully acknowledging that our lives are comprised to a great extent not of what we choose to do, but of what happens to us. To see each other not just as choosing agents (who choose, among other things, what they desire to remember), but as human beings who must cope with circumstances, accidents, bad (moral) luck, unfortunate choices, and so on. Our traumatic, unwanted memories keep us in the same moral universe, working with the same flawed moral understandings of our world and of each other. The alternative—the chosen forgetting, the selective dampened memories—not only have the capacity to erase those bonds, but quite literally, like *Eternal Sunshine's* Joel Barish and Clementine Kruczynski, to mutually write others out of our lives and re-cast them as strangers to whom we have never made ourselves vulnerable. What I mean is this: by deliberately forgetting, I am rewriting not only the story of my life, but the story of all those other lives whose social, political, and personal realities have intersected with my own in ways that *mattered to them*. But because my actions are exclusively self-regarding, these others, as they become epistemically vague, also grow increasingly morally-insignificant—*I have chosen to write them out*. And thus, the shared moral universe in which we move through each other's identity-constituting stories shrinks—as do our capacities for mutual moral, political, and other undertakings.

But of course, I am arguing for techno-incrementalism, and against techno-optimism—not for a kind of a neo-Luddite avoidance of technological innovation. My worries about the othering of the self and of others must be balanced against concerns about human flourishing, which PTSD and other responses to traumatic experiences often make very challenging, if not impossible. To force an assault survivor to remain locked up with her debilitating memories, or to leave a veteran with his psychological wounds, would not only be a kind of political othering in the sense of abandoning the vulnerable as socially unwanted and uncared-for outcasts—it would clearly also be cruel. There are cases, many of them, where worries about memory modification will be overruled by the possibilities for healthier, more fulfilling, more endurable, lives. Thus what my call for techno-incrementalism encourages is a kind of a balancing act between two ambiguities—between cases where memory modification would be potentially destructive, and those where it offers a glimmer of hope. But we must consider the consequences before we leap, not just for the individual patient, but for all those who might be affected by the choice to forget—and that number, it seems to me, is significant enough to give us pause.

3. Avatars, Virtual Spaces, and Being "Otherwise"

In choosing to address virtual reality as a kind of (bio)technology, I am making two claims: first, that these technologies are becoming intimately connected with human embodiment; second, that they are powerful shapers of the social relations that shape this embodiment. My focus is both on virtual reality more generally, and on a specific application of virtual spaces and avatars called *Second Life* (hereinafter SL). I must note that although virtual reality is gaining in popularity, SL is not. My reasons for addressing SL have more to do with its initial promise and its resulting failures, for in those failures, we can glimpse some of the reasons for worries about virtual technologies as means of (bio)enhancement that is meant to allow human beings to experience, and perhaps even understand, life in a way that is not otherwise possible. I begin by addressing the broader uses, and narratives about, virtual reality.

3.1 Virtual Reality as "Empathy Machine"

What I mean by "virtual reality" (VR) is a kind of phenomenological experience, mediated by technologies that give the user the illusion of being in, and engaging with, environments that range from the ordinary to the unusual to the imagined to the seemingly impossible. These spaces, accessed through VR glasses or other technology that the user usually wears over his or her eyes, have offered entertainment, "travel" to distant places, and, as Chris Milk, founder and CEO of the VR company Within, argued in his 2015 TED talk, the "ultimate empathy machine" (Kang 2017).

What he meant was echoed in several recent efforts to employ VR technologies as a means of helping people (mostly from the Global North) empathize with those situated differently from themselves by virtually "experiencing" what it is like to be otherwise. For example, in a recent New York City fundraiser by the International Rescue Committee, headsets offered experiences of a refugee camp in Lebanon (Bloom 2017). Stanford University VR simulations attempt to offer the experience of getting evicted and becoming a homeless, disabled, schizophrenic (Bloom 2017). Moreover,

> Clouds Over Sidra, the U.N.'s first virtual reality documentary, follows a young Syrian refugee named Sidra who lives in a refugee camp in Jordan [. . .]The user is able to look out across the desert in all directions, up to the sun, and down to the sand below her. If the user looks directly down, she doesn't see her feet or the base of the camera, but a collection of footprints, whose tracks span across the desert around her. She has no feet in this world, no physical form at all, but the marks of multiple feet are beneath her. Disembodied, she occupies the point at which all the footprints meet (Hamilton 2017).

In promoting the "transformative" effects of VR experiences, Mark Zuckerberg, using an Oculus Rift and Facebook's virtual reality platform, Facebook Spaces, "traveled" to, among other places, Puerto Rico after the destruction of Hurricane Maria, using the purported immediacy of the experience to argue for aid relief (Kastrenakes 2017). Building on the promises of phenomenological access to the other (or the "othered"), the *Frontline* VR documentary "After Solitary" (Mucciolo and Herrman 2017) focused on the life of Kenny Moore, who was found guilty of aggravated assault, burglary, and theft as a teenager, and spent twenty years in and out of prison—five of them in solitary confinement (Mucciolo and Herrman 2017). The viewer is sometimes left alone in the cell, her Oculus Rift attempting to place her in the middle of the suffering that is part and parcel of the carceral experience, and, in the end, elicit something like empathy. Indeed, *Frontline*, in advertising the program, asks: "What's it really like to be locked up in solitary confinement?" (Kang 2017). But has the "empathy machine" experiment worked? Has virtual reality, in its emerging guises, served as not simply a means of new entertainment, but what its proponents argued was its greater purpose?

Not so far—or, at least not to any extent where the VR-as-access-to-the-other claim can be a justification for its continual development. At best, VR might generate certain sympathies—but this is far from being the empathy-generator that its proponents have suggested (Ramirez 2018), and can greatly depend on the extent to which individuals already identify with those whose experiences or persons the VR technology is attempting to reify (Avenanti et al. 2010). While I think that it is not a dispositive argument against the further development of VR—after all, entertainment is an important part of the human experience, and a number of VR creators are genuinely interested in its more altruistic uses—I would argue that no clever technological tools can force empathy. In other words, one cannot make another care against their will.

The reasons for this are complex. We might ask why a VR headset, and the world into which it is designed to project the viewer, is a better empathy pump then the "meatspace" in which we ordinarily find ourselves. It does not seem like the advanced graphics and the you-are-there effects would be "more real" than the world that surrounds us on a daily basis. So the difference has to lie elsewhere—perhaps in the ability of VR to transport us to places we would not otherwise be able to access, whether physical locations, experiences, or the phenomenologies of distant others. Yet the very concept of empathy creation through VR is a process of othering: the unspoken message is that the experiences of some are so vastly different from one's own that one could only find common ground with them by virtually inhabiting their (virtual) spaces. This presumption of otherwise unshareable worlds and unspeakable differences has three consequences: first, an "othering" of people and experiences too other-worldly and bizarre to be grasped without technologically-assisted means; second, a reduction of the possibility of being otherwise to a technologically-mediated—and usually entertaining—experience; and third, a "self-othering": the loss of the capacity to live outside of non-virtual worlds.

The results of the first and second consequence can be seen in the casual ease with which the makers, and the users, of VR technologies create and consume encounters with distant "others": VR is sold and purchased with the promise of the experience of *being there*. But here is the worry: VR does not help one to empathetically enter the world of another—one might think that one is doing so because of the "realness" and newness of one's virtual surroundings, but what one is, at best, is a tourist. And tourists do not often pause long enough, or deeply enough, to inhabit an alien environment, a stranger's life. Instead, they merely pass through, remarking on the greatness, the awfulness, the beauty, or the ugliness of a place, of a person or a people:

> The problem is that these experiences aren't fundamentally about the immediate physical environments. The awfulness of the refugee experience isn't about the sights and sounds of a refugee camp; it has more to do with the fear and anxiety of having to escape your country and relocate yourself in a strange land. Homeless people are often physically ill, sometimes mentally ill, with real anxieties about their future. You can't tap into that feeling by putting a helmet on your head. Nobody thinks that going downtown without your wallet will make you appreciate poverty—why should these simulations do any better? One specific limitation of VR involves safety and control. During the debates over the interrogation practices of the United States during the Iraq war, some adventurous journalists and public figures asked to be waterboarded, to see what it was like. They typically reported that it was awful. But in fact their experience fell far short of how terrible actual waterboarding is, because part of what makes waterboarding so bad is that you get it when you don't want it, by people who won't stop when you ask them to.
>
> (Bloom 2017)

Further, Jim Blascovich and Jeremy Bailensen note that "[. . .]virtual realities easily satisfy[. . .]social needs and drives—sometimes [they are] so satisfying that addicted users will withdraw physically from society" (Kim 2015). For instance, in South Korea, one of the most wired nations in the world, camps have been set up to reintroduce technology-addicted young people to life and communication in the analog world (Fifield 2016). Moreover, while the Japanese "hikikomori" phenomenon of the fully-socially-withdrawn, and often technology-dependent, young-to-middle age adult was once culturally limited, it is now becoming a worldwide concern as advanced technologies are beginning to supplant unmediated experiences and connections with the real world (Adamski 2018). And even though escape from the daily pressures of life—and from conditions such as anxiety, depression, PTSD, and others—is not unimportant, the potential for total VR-centered rejection of the external world is not insignificant. As VR becomes more affordable and more accessible, there might very well be an increase in "the size of the population for whom more highly immersive perceptual and psychological experiences are available," and or whom the non-VR experiences become increasingly secondary and peripheral to their lives (Kim 2015).

3.2 Virtual Identities as (Failed?) World-Traveling

The philosopher Maria Lugones has argued that one is not merely an immigrant, or an elderly person, or a patient, but a multifaceted locus of identities, an inhabitant of many worlds—worlds between which one can travel (Lugones 1987). I can be a focused professional in one world, a chatty friend in another, and a devoted family member in yet another, and my personal identity is comprised of them all. One might even argue that without much world-traveling, one's idea of oneself becomes shallow, fixed, two-dimensional, and without the capacity to understand differently-situated others. But while some kinds of world-traveling have to do with occupying different roles in socio-physical spaces, others are meant to cross the boundaries not just of selves, but of the physical-virtual schism itself.

But not all experiments in world-traveling succeed. And so, I turn to a specific case of a largely failed technological attempt at simulated world-traveling as an example of what happens when we cast ourselves, unprepared, into new domains. What I have in mind is a virtual space from the early-to-mid 2000s called *Second Life*. I have two reasons for addressing it specifically: First, while *Second Life* no longer enjoys the popularity it had five-to-ten years ago, it is still in existence, more recently as *Sansar*, a VR platform. Second, the techno-optimism and anticipation that surrounded its release—and its eventual (general) failure to deliver the hoped-for experiences—supports the modification of such optimism to what I call techno-incrementalism, which takes seriously the othering, alienating aspects of our evolving techno-selves.

As I have discussed in an earlier paper focused entirely on its nature, intent, and consequences, *Second Life*, launched in 2003 by Linden Lab, is intentionally designed as an environment that is to some extent a blank canvas, with a number of rules, to be populated, imagined, and constructed by its users. Largely inspired by the hyperreality of the "Metaverse" environments of Neal Stephenson's novel *Snow Crash*, SL was intended to become a place where anyone could create an avatar body, as well as the space that the body occupies, limited mainly by one's imagination and technical abilities (Boellstorff 2008; Gotlib 2014).

In SL, environments and bodies are not flesh-and-bone. They are ordinary or fantastical, heroes and heroines, or else creatures who do not share any of our human corporeal shapes at all. And while VR games are usually limited not only by avatar choices and the teleology of the game itself, SL's avatars have no pre-established limits, save for those imposed by the technology itself: one can switch between species, homes, pursuits—or do nothing at all.

Without pre-existing narratives of self or space, anthropologist Tom Boellstorff describes SL as a kind of self-directed, chaotic world:

> [a] man spends his days as a tiny chipmunk, elf, or voluptuous woman. Another lives as a child and two other persons agree to be his virtual parents. Two "real"-life sisters living hundreds of miles apart meet every day to play games together or shop for new shoes for their avatars. The person making the shoes has quit his "real"-life job because he is making over five thousand U.S. dollars a month from the sale of virtual clothing. A group of Christians pray together at a church; nearby another group of persons engages in a virtual orgy ... Not far away a newsstand provides copies of a virtual newspaper with ten reporters on staff; it includes advertisements for a "real"-world car company, a virtual university offering classes, a fishing tournament, and a spaceflight museum with replicas of rockets and satellites.
>
> (Boellstorff 2008, 17)

This virtual life feels quite real to its participants. As one resident noted, "this is how I see myself on the inside" (Boellstorff 2008, 134). In fact, Raffaele Rodogno has argued that "online activities may, in different ways, affect our offline personal identity [. . .] [T]he more important online activities become [. . .] the more we can suspect that any self-narrative [one] would recount would include events that occur within them[. . .] [O] ur interpretation of ourselves is constitutive of who we are," and thus our "identity will significantly be constituted by[. . .]online life ... " (Rodogno 2012). Marya Schechtman, noting the findings of ethnographer Annette Markham, suggests that virtual embodiment is in fact a new kind of identity-constituting reality:

> Markham reports that she "found reason to destabilize a traditional idea that the experience of reality is grounded in the physical, embodied world." To her surprise, she says, the residents she engaged with the question "What is really real?" told her "this question was of little relevance to them; rather everything that is experienced is real ... [and] our virtual relationships are just as real as our rl [real life] ones."
>
> (Schechtman 2012, 331)

So where is the problem? It lies in two distinct, yet related, aspects of SL: First, even though its immersive, non-game-focused environment might appear to promise otherwise, empathy toward others is not only rare, but, in fact, too often the opposite of what actually takes place. Stereotypes, prejudices, and other real-world bad habits take root: Those with female-presenting avatars strive to be young, white, and thin—and as scantily-clad as possible; those with male-presenting avatars often choose the more typically macho, muscular, aggressive-acting identities. Indeed, once we move beyond the novelty of technological trappings, high-tech versions of the familiar narratives of beauty, gender roles, ability, and social hierarchy tend to not only repeat themselves, but dominate the virtual world of ostensibly limitless possibilities (see Belamire 2016). And thus instead of transcending the corporeal world's oppressive storylines and social dictates, thereby gaining new, more empathy-grounded insights into ourselves and others—SL seems to cement the worst of what we already do.

Second, SL, while promising to enhance our practices of racial tolerance by offering not only the opportunity to present as someone of another ethnic identity, but to bring together people from all over the world in a place where all are equal by virtue of being "virtual," collapsed into an often-vicious version of virtually-embodied, powerfully othering, racisms. As Boellstorff noted:

> some residents who tried wearing nonwhite skins reported racist responses, including friends who stopped answering [instant messages] and statements that nonwhite persons were invading *Second Life*. It is not surprising that some residents who were nonwhite in the actual world engaged in forms of racial passing, so that at least one of their avatar embodiments was white
>
> (Boellstorff 2008, 145).

Not only were non-white avatars othered, but individuals who were non-white in the non-virtual world felt alienated enough to hide who they were—forced, in a sense, to "pass," to fit in to what was becoming a space where it might be all right to appear as an octopus or a non-living entity, but not as a black or brown individual. Moreover, instead of a newfound freedom to experiment with both embodiment and identity, a hardening of group privilege took place. As Lisa Nakamura argues, participants engaged in "identity tourism"—a kind of a journeying of the already privileged through the experience of being an "other," whether that other is a racial or other kind of minority. In its focus on the "mysterious Asian other," this is an act of both Orientalism and exotification. In its attempted play-acting of a minority identity, it is an example of supreme sociopolitical privilege. For instance,

> players who choose to perform this type of racial play are almost always white, and their appropriation of stereotyped male Asiatic samurai figures allows them to indulge in a dream of crossing over racial boundaries temporarily and recreationally . . . Tourism is a particularly apt metaphor to describe the activity of racial identity appropriation, or "passing" in cyberspace . . . [I]dentity tourism . . . allows a player to appropriate an Asian racial identity without any of the risks associated with being a racial minority in real life
>
> (Nakamura 2000).

What is a game for some is fundamentally othering for those who, on the one hand, are punished for presenting as their own (minority) identities, and on the other, are used as an amusing "skin" for the consequence-free entertainment of privileged others. Quite similar to the "tourism" of VR-based experiences critiqued earlier, the SL-kind of immersion is, in some sense, more damaging. Instead of even pretending to bring the participant closer to unknown phenomenologies, it offers a playground where the dominant can continue dominating and making the other liminal—making them not politically or morally relevant to the dominant society—while themselves remaining uncoupled from the possibility of public shame. Indeed, identities become something akin to a game that has very little to do with either deepening or broadening our universe or enhancing our capacity to engage in more inclusive practices of moral understandings. And even though SL and SL-type experiences do have the potential to become the kind of liberating, ongoing experiment in world-traveling that its creators had envisioned—after all, SL and other virtual environments have shown to offer some liberatory potential for people with disabilities, as well as for some kinds of therapies, and so on (Boellstorff 2008)—both SL's relatively fast loss of popularity, and its troubling consequences for the already-vulnerable, suggest a cautious approach to similar endeavors. What we are trying to achieve, how we are trying to achieve it, and what some of the intended and unintended consequences might be, ought to figure in both the decision to proceed and the justifications for so doing. But like with other virtual world experiments and enhancements, their growth and expansion should be tempered by a realization of their power to shape, define, and destroy personal and group identities—and their subsequent potential to further disempower the already-vulnerable, to further other the already liminal.

4. Have We Got a Robot for You: The Biotechnologies of Aging

Finally, I would like to examine a third biotechnological trend that merits some worry: the growing adaptation of robotics to the care of the elderly (Sparrow and Sparrow 2006; Sorell and Draper 2014). Once again, my point is not that such technological innovation ought to be halted, nor even that the arguments against it outweigh those that support its development. Rather I suggest that we ought to pay attention to some of the ways these technologies may misunderstand, other, or simply get wrong the care relationships into which they are placed, and which they are, ostensibly, meant to improve. Specifically, I argue that robotics-based care for the elderly opens up the possibility for several kinds of losses, including a loss of human contact and connection—indeed, it becomes a kind of othering, leading to a greater liminality of the aged.

4.1 Robots Who Care

Caregiving is often emotionally difficult, physically demanding, intimate work. It asks the care-giver to be at once vigilant and gentle, understanding and firm, all the while remaining in a personal as well as a professional relationship with the patient. It is also work that is routinely not well-paid, often psychologically and physically punishing, mostly done by women and immigrants, and not something that most people with the means to outsource it have the desire to do. Perhaps unsurprisingly, there is an ongoing shortage of care workers for the aging in the Global North. To add to these challenges to good care, too many people find themselves in care facilities with very little contact with their family and friends, further isolating them from the kind of connections, interactions, mental stimulation, and emotional support that human flourishing requires, and pushing them, further into loneliness, depression, and physical and psychological decline (Resnick 2017). There is perhaps no more powerful example of the extreme consequences of such isolation than the COVID-19 pandemic of 2019–2020, when the elderly not only suffered in care facilities that were understaffed (sometimes, to the point of near-abandonment), under-resourced, and overrun with illness, but as many afflicted with coronavirus, they died alone, as well. As Mary Pipher (2000) notes, the old often live, even if figuratively, in another country (Gotlib 2014).

In response, a number of countries—mostly notably Japan—are beginning to introduce caregiving robots to take over for overworked or absent human care providers (Turkle et al. 2006; Vandemeulebroucke et al. 2018). Indeed,

> [. . .] in Japan, where robots are considered "iyashi," or healing, the health ministry began a program designed to meet work-force shortages and help prevent injuries by promoting nursing-care robots that assist with lifting and moving patients. A consortium of European companies, universities and research institutions collaborated on Mobiserv, a project that developed a touch-screen-toting, humanoid-looking "social companion" robot that offers reminders about appointments and medications and encourages social activity, healthy eating and exercise[. . .]Researchers in the United States are developing robot-caregiver prototypes as well, but we have been slower to move in this direction. Already [. . .] robots [. . .] assist in surgery and very basic "walking" robots that deliver medications and other supplies in hospitals
>
> (Aronson 2014).

So far, so good—after all, helping aging individuals to move, and reminding them to engage with the world (or even to take their medication) seems to have no drawbacks—especially given the growing number of care-receivers, and the understaffed facilities staffed by overworked care providers. Even a techno-pessimist might have a difficult time objecting to these purely assistive robots. However, the significant distinction to watch for is the move from task-oriented "assistant" technologies to robots that are meant to take over some of the "softer," interpersonal, human relationships that neither families of the aging, nor care-giving staff, are willing or able to meet. For instance,

> [. . .] the Care-O-bot being developed in Germany, a funky looking wheeled robot with arms and a round screen as a head. It might be a bit more human-like, as with the Pepper personal robot from Softbank Robotics, with its cartoonish head and a screen on its chest. Pepper's claim to fame is its ability to "perceive emotions." ("Pepper" is the first humanoid robot capable of recognizing the principal human emotions and adapting his behavior to the mood of his interlocutor, according to the company's website).
>
> (Outing 2017)

I suggest that the moral worry can be located less in the fact of the advancing technology, but in the technological substitutes that are designed for the "emotional," rather than merely physical, human labor of caregiving. As Sherry Turkle argues in a 2012 TED talk, "I believe that this breaks the compact between generations. These machines do not understand us. They pretend to understand ... To me, it lessens us. It is an inappropriate use of technology. With the right social priorities, we get more people in jobs for seniors. Asking robots to do this job is asking more of machines and less of each other" (Outing 2017).

While I do not agree with what I view as Turkle's strong techno-pessimism, I do think that we ought to proceed carefully with regard to the kinds of care robots that do not merely assist human caregivers, but purport to serve as substitutes for both their physical and emotional roles. The reasons for my call for pause have to do with the resulting othering of the elderly— he creation of technologically-facilitated liminality of the already-vulnerable, born of powerful techno-optimism and socio-economic desire for efficiency and increased productivity. By "liminality," what I mean is that those burdened by illness, age, and related issues already find themselves in-between: in-between health and wellness; freedom and confinement; being a part of something and isolation, and so on. When this kind of "out-of-place" vulnerability is further compounded by the loss of human connections and replaced with automata, the liminal elderly become the "other"— a body whose needs can be met via any means of *delivery of services* rather than a human being whose personhood calls for *care*. Thus, the master narratives of what "they need" are now not merely bureaucratic boilerplate, but reified in the methods of delivery itself.

4.2 Care as Product

As noted earlier, the elderly, whether they live in their own homes or in a care facility, are already marginalized by physical isolation, loneliness, and distance from family and friends. And in part due to this alone-ness, they are vulnerable both physically and emotionally. Living alone not only tends to result in depression, but has been correlated with dementia (Fratiglioni et al. 2000), and specifically with Alzheimer's disease (Wilson et al. 2007). But their vulnerability is not just about susceptibility to illness: without the support of close others nearby, they are often socially, economically, and psychologically liminal. If in their own home, they face many hours without human companionship; if in a residential facility, their freedom to move, to participate in activities, and even to socialize is often tightly controlled by an impersonal institution, too often staffed by exhausted, overworked, and underpaid staff. So why would the assistive technology not help—especially if its intended uses include emotional support?

The first worry is that care becomes a product, and care-giving becomes a delivery mechanism, where the quality of what is delivered, whether medication, food, or "love," matters less than the efficiency and speed at which it is carried out. A reminder to take a stroll equals a reminder to watch one's favorite television program equals a reminder that one is loved. Meanwhile, human contact, for all of its expense, complication, and messiness, is gradually eliminated altogether, becoming a "machinery of care," and further othering and making socially liminal an already-vulnerable population.

The second worry has to do with the objectification of the elderly and the nature of care itself. As I argued in a previous section, identities are largely narrative, and are created by a multiplicity of stories, told by, and about, us from first, second, and third-person perspectives (Gotlib 2015). When one becomes dependent on nursing care, one's life story, and thus one's identity, can be significantly underwritten by one's relationships with those on whom one depends, whether family members, a friend, a nurse, or any other human care-giver. It matters very much if said care is offered with compassion, kindness, even humor, or if it is merely grudgingly doled out, barely there, tethered only by a sense of duty or economic necessity. In the former case, one's humanity, one's personhood is upheld; one's identity as someone requiring help woven into the larger narrative of who one is—it is made coherent, legitimate, bearable. As Hilde Lindemann puts it, one is "held in personhood" (Lindemann 2014). In the latter, one may be objectified, painted as a burden, as not the person one once was, and thus the narrative of one's life may be broken, damaged, made unrecognizable to the care-recipient and the care-giver alike.

Yet in both of these cases, the care-giver, regardless of how engaged or withdrawn, is human. What happens when the care-giver is a machine—especially when it is a machine that is responsible for the physical and psychological well-being of elderly persons—is that "care" becomes re-defined as merely another service, and its subject is objectified as something-to-be-serviced. Put another way, the possibility of integrating the new caring relationship between the patient and the care-giver into their respective identities is foreclosed: the robot cannot meaningfully add to a patient's story, and it itself has no concerns about its own identity, diachronic or otherwise. The story of who the cared-for *is* is reduced to an interaction with an object, albeit one built to resemble an agent. The cared-for, therefore, might go through three stages of coping with the loss of human companionship: (1) she first might be deceived into thinking that she has a relationship with the robotic care-giver; (2) she might then realize that such a relationship is impossible, and is faced with the resulting loneliness and liminality; and (3) she is further othered, as a result, by a technology ostensibly created to be of assistance. Soon, the robotic becomes the default for those who do not possess the resources to insist on human care. And just as their objectification becomes a part of their identity, so their stories about who they are become unrecognizable, broken, or silenced—or disappear altogether.

5. Conclusion: A Plea for Techno-Incrementalism

In this chapter, I have attempted to argue against the growing acceptance of techno-optimism—a view that future technologies are the panacea that will, once and for all, solve our current (and future) problems. However, neither am I advocating for a kind of despondent techno-pessimism, put forth by those worried about technological progress' dehumanizing, destabilizing, and generally harmful effects on both individuals and societies. Instead, I conclude by suggesting that we adopt the stance of techno-incrementalism: a careful, deliberate approach to biotechnologies that neither rushes headlong into "progress," nor refuses to engage with emerging technologies out of (however well-founded) fears.

What this means is that we take the middle ground, centering the vulnerable person rather than either our enthusiasm or fears regarding technological progress. This means that we address the suffering born of traumatic memories by first acknowledging that not everything unpleasant ought to be forgotten, and second, by considering that selective memory manipulation might be dangerously close to transforming a disfavored narrative of our lives into a more pleasing, yet more fictitious and othering one. This note of caution does not mean that we cruelly condemn victims of violence, soldiers, and survivors of other potentially PTSD-inducing events to suffering when a biotechnological discovery might make it a bit more tolerable, and not entirely debilitating. It merely means that we more deeply and more carefully consider the consequences, both political and personal, short-term as well as long-term, of our choices (Romm 2014).

Virtual reality, with all of its attendant worlds and avatars, might indeed someday become an "empathy machine"—but not if we embrace it for its novelty and excitement without considering the kinds of social and personal identity-centered worries it creates. The techno-incrementalist moves slowly, giving sufficient time and attention to questions of power differentials and oppressions that are no mere abstractions made possible by virtual presence. The incrementalist asks for evidence that a particular virtual technology lives up to its promises and responsibilities—or at least does not wholly fail to do so.

Finally, a techno-incrementalist understands that the complexities and vulnerabilities of aging, while assisted by emerging technologies, are not at the same time solved through their use. Neither loneliness, nor isolation, nor powerlessness, nor dependency are addressed by replacing human contact, care, and attention by robots—no matter how user-friendly, cuddly, or life-like. What is needed is a re-assessment of the changing needs of an aging population that first and foremost asks what this population wants—not what would be most socially and politically expedient, least expensive, and least demanding of the rest of us. These considerations must include, among other things, large political initiatives to attract, and retain, well-educated, willing, and capable caregivers—initiatives that must focus on fair and generous compensation and benefits, good working conditions, and greater respect for this challenging and necessary work. Only then might the incrementalist turn to assistive technologies, such as helper robots, to fill in whatever human caregivers cannot provide.

My intent is merely to suggest that we slow down, take a deep breath, and consider where our recent mood of techno-optimism is taking us, especially when it comes to the care for the identities of vulnerable others. Perhaps the techno-optimist is too wedded to the notion of a particular kind of progress that fails to be accountable for its traumatizing costs; perhaps he deliberately looks past the more socially-grounded and less technology-dependent solutions, such as increased mental health services instead of treatments to forget the trauma altogether; direct, empathy-creating contact with distant others; better-compensated and respected elder care workers, offering non-othering, non-liminal-making care, and so on. Indeed, it seems that a slower, more deliberate, more careful approach to (bio)technological evolution—one that takes seriously questions of political power, of personal identity, and of the othering born of technological innovation—would neither slow down progress in a way that is detrimental to human flourishing, nor would irreparably damage our complicated relationship to technology. Instead, it just might make us face our creations with a greater sense of responsibility to ourselves, and to each other.

References

Adamski, Dawid. 2018. "The Influence of New Technologies on the Social Withdrawal (Hikikomori Syndrome) among Developed Communities, Including Poland." *Social Communication* 1, no. 17: 58–63.

Aronson, Louise. 2014. "The Future of Robot Caregivers." *The New York Times*, July 2014. https://www.nytimes.com/2014/07/20/opinion/sunday/the-future-of-robot-caregivers.html

Avenanti A., Sirigu A., and Aglioti S.M. 2010. "Racial bias reduces empathic sensorimotor resonance with other-race pain." *Current Biology* 8, no. 20(11) (June): 1018–1022.

Beckford, Martin. 2011. "Elderly People Isolated by Technological Advances." *The Telegraph*, November 2011. https://www.telegraph.co.uk/news/health/elder/8867767/Elderly-people-isolated-by-technological-advances.html

Belamire, Jordan. 2016. "My First Virtual Reality Groping." *Medium*, October 2016. https://medium.com/athena-talks/my-first-virtual-reality-sexual-assault-2330410b62ee

Bell, J. 2008. "Propranolol, Post-Traumatic Stress Disorder and Narrative Identity." *Journal of Medical Ethics* 34, no 11 (Nov.): 23.

Birey, Fikri. 2014. "Memories Can Be Edited." *Scientific American*, May 2014. https://www.scientificamerican.com/article/memories-can-be-edited/

Bloom, Paul. 2017. "It's Ridiculous to Use Virtual Reality to Empathize with Refugees." *The Atlantic*, February 2017. https://www.theatlantic.com/technology/archive/2017/02/virtual-reality-wont-make-you-more-empathetic/515511/

Boellstorff, T. 2008. *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*. Princeton: Princeton University Press.

Carey, Benedict. 2009. "Brain Power: Brain Researchers Open Door to Editing Memory." *New York Times*, April 6, 2009. http://www.nytimes.com/2009/04/06/health/research/06brain.html).

Chadwick, Alex. 2004. "Analysis: Concepts in Memory-Loss Movies Not So Far-Fetched." *NPR*, March 23, 2004.

Chandler, Jennifer A., Alexandra Mogyoros, Tristana Martin Rubio, and Eric Racine. 2013. "Another Look at the Legal and Ethical Consequences of Pharmacological Memory Dampening: The Case of Sexual Assault." *Journal of Law, Medicine & Ethics* (Winter): 859–871.

Charon, R. and Montello, M. 2002. *Stories Matter: The Role of Narrative in Medical Ethics*. New York: Brunner- Routledge.

Cross, Katherine. 2016. "Online Harm Is Real." *Slate*, November 2016. https://slate.com/technology/2016/11/sexual-harassment-in-virtual-reality-is-real.html

Cross, Katherine. 2016. "The Rise of Virtual Sexual Assault and Cyber-Groping in Virtual Reality Gaming." *Salon*, November 2016. https://www.salon.com/2016/11/11/the-rise-of-virtual-sexual-assault-and-cyber-groping-in-video-games_partner/

DeGrazia, David. 2005. "Enhancement Technologies and Human Identity." *Journal of Medicine and Philosophy* 30: 261–283.

Donovan, Elise. 2010. "Propranolol Use in the Prevention and Treatment of Posttraumatic Stress Disorder in Military Veterans: Forgetting Therapy Revisited." *Perspectives in Biology and Medicine* 53, no. 1, (Winter): 61–74.

Elliott, Carl. 2011. "Enhancement Technologies and the Modern Self." Journal of Medicine and Philosophy 36: 364–374.

Erler, Alexandre. "Does Memory Modification Threaten Our Authenticity?" Neuroethics 4 (2011): 235–249. doi:10.1007/s12152-010-9090-4

Evers, Kathinka. 2007. "Perspectives on Memory Manipulation: Using Beta-Blockers to Cure Post-Traumatic Stress Disorder." *Cambridge Quarterly of Healthcare Ethics* 16, no. 2 (Spring): 138–46.

Fifield, Anna. 2016. "In South Korea, a Rehab Camp for Internet-Addicted Teenagers." *The Washington Post*, January, 2016. https://www.washingtonpost.com/world/asia_pacific/in-south-korea-a-rehab-camp-for-internet-addicted-teenagers/2016/01/24/9c143ab4-b965-11e5-85cd-5ad59bc19432_story.html

Fratiglioni, Laura, Hui-Xin Wang, Kjerstin Ericsson, Margaret Maytan, and Bengt Winblad. 2000. "Influence of social network on occurrence of dementia: A community-based longitudinal study." *Lancet* 355: 1315–1319. doi:10.1016/S0140-6736(00)02113-9.

Glannon, W. 2006. "Psychopharmacology and Memory." *Journal of Medical Ethics* 32, no. 2 (Feb.): 74–78.

Gondry, Michel, Steve Golin, Anthony Bregman, Charlie Kaufman, Pierre Bismuth, Jim Carrey, Kate Winslet, et al. 2004. *Eternal Sunshine of the Spotless Mind*. https://www.imdb.com/title/tt0338013/

Gotlib, Anna. 2014. "Girl, Pixelated: Narrative Identity, Virtual Embodiment, and Second Life," *Humana Mente: Journal of Philosophical Studies* 26 (May): 153–178.

Gotlib, Anna. 2014. "Intergenerational Justice and Health Care: A Case for Interdependence." *International Journal of Feminist Approaches to Bioethics* 7 (1): 142–168.

Gotlib, Anna. 2015. "Beyond the Trolley Problem: Narrative Pedagogy in the Philosophy Classroom," In *Feminist Pedagogy in Higher Education: Critical Theory and Practice*, edited by Tracy Penny Light, Jane Nicholas, and Renée Bondy. Waterloo, Ontario: Wilfrid Laurier University Press.

Gräff, J., Joseph, N. F., Horn, M. E., et al. 2014. "Epigenetic Priming of Memory Updating during Reconsolidation to Attenuate Remote Fear Memories." *Cell* 16; no. 156(1–2) (Jan.): 261–276.

Hamilton, Kathryn. 2017. "Voyeur Reality." *The New Inquiry*, February 2017. https://thenewinquiry.com/voyeur-reality/

Han, Jin-Hee, Steven A. Kushner, Adelaide P. Yiu, et al. 2009. "Selective Erasure of a Fear Memory." *Science* 13 (Mar.): 1492–1496.

Hurley, Elisa A. 2007. “The Moral Costs of Prophylactic Propranolol.” *The American Journal of Bioethics* 7 (9): 35–36.

Kang, Inkoo. 2017. “Oculus Whiffed.” *Slate*, November 2017. https://slate.com/technology/2017/11/virtual-reality-is-failing-at-empathy-its-biggest-promise.html#

Kastrenakes, Jacob. 2017. “A Cartoon Mark Zuckerberg Toured Hurricane-Struck Puerto Rico in Virtual Reality.” *The Verge*, October 2017. https://www.theverge.com/2017/10/9/16450346/zuckerberg-facebook-spaces-puerto-rico-virtual-reality-hurricane

Kenwright, Ben. 2018. “Virtual Reality: Ethical Challenges and Dangers,” *IEEE Technology and Society Magazine* 37 (4): 20–25.

Kim, Monica. 2015. “The Good and the Bad of Escaping to Virtual Reality.” *The Atlantic*, February 2015. https://www.theatlantic.com/health/archive/2015/02/the-good-and-the-bad-of-escaping-to-virtual-reality/385134/

Liao, Matthew S, and Anders Sandberg. 2008. “The Normativity of Memory Modification.” *Neuroethics* 1: 85–99 doi:10.1007/s12152-008-9009-5

Lindemann, Hilde. 2014. *Holding and Letting Go: The Social Practice of Personal Identities*. New York: Oxford University Press.

Lindemann Nelson, Hilde. 2001. *Damaged Identities, Narrative Repair*. Ithaca: Cornell University Press.

Lu, Stacy. 2015. “Erasing Bad Memories.” *Monitor on Psychology* 46, no. 2 (February). https://www.apa.org/monitor/2015/02/bad-memories.aspx

Lugones, Marıa. 1987. “Playfulness, ‘World’ Traveling, and Loving Perception.” *Hypatia* 2 (2): 3–19.

Miller, Greg. 2004. “Learning to Forget.” *Science, New Series* 304, no. 5667 (Apr.): 34–36.

Mucciolo, Lauren and Cassandra Herrman. 2017. *After Solitary*. Filmed 2017. https://www.pbs.org/wgbh/frontline/article/after-solitary/

Nakamura, L. 2000. “Race in/for Cyberspace: Identity Tourism and Racial Passing on the Internet.” Accessed, May 4, 2020. http://www.humanities.uci.edu/mposter/syllabi/readings/nakamura.html.

Outing, Steve. 2017. “Is There a Robot ‘Friend’ in Your Future?” *Forbes*, October 2017. https://www.forbes.com/sites/nextavenue/2017/10/04/is-there-a-robot-friend-in-your-future/#768ad708516f

Parens, Erik. 2010. “The Ethics of Memory Blunting: Some Initial Thoughts.” *Frontiers in Behavioral Neuroscience* 4, no. 190 (Dec.). doi:10.3389/fnbeh.2010.00190

Pipher, Mary. 2000. *Another Country: Navigating the Emotional Terrain of Our Elders*. New York: Riverhead Books.

President’s Council on Bioethics (U.S.), and Leon Kass. 2003. *Beyond therapy: biotechnology and the pursuit of happiness*. Washington, D.C.: President’s Council on Bioethics.

Ramirez, Erick. 2018. “It’s Dangerous to Think Virtual Reality Is an Empathy Machine.” *Aeon*, October 2018. https://aeon.co/ideas/its-dangerous-to-think-virtual-reality-is-an-empathy-machine

Resnick, Brian. 2017. “Loneliness Actually Hurts Us on a Cellular Level,” *Vox*, January 2017. https://www.vox.com/science-and-health/2017/1/30/14219498/loneliness-hurts

Rodogno, R. 2012. “Personal Identity Online.” *Philosophy and Technology* 25 (3): 309–328.

Romm, Cari. 2014. “Changing Memories to Treat PTSD.” *The Atlantic*, August 2014. https://www.theatlantic.com/health/archive/2014/08/changing-memories-to-treat-ptsd/379223/

Schechtman, Marya. 2007. “Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View.” *Royal Institute of Philosophy Supplement* 60: 155–178.

Schechtman, Marya. 2012. “The Story of My (Second) Life: Virtual Worlds and Narrative Identity.” *Philosophy and Technology* 25 (3): 329–343.

Schwarzenegger, Arnold, Rachel Ticotin, Sharon Stone, Michael Ironside, Ronny Cox, Paul Verhoeven, Ronald Shusett, and Philip K. Dick. 2001. *Total Recall*. Santa Monica, CA: Artisan Entertainment.

Smith, Aaron. 2014. “U.S. Views of Technology and the Future Science in the Next 50 Years.” *Pew Research Center*, April 17, 2014. https://www.pewresearch.org/internet/2014/04/17/us-views-of-technology-and-the-future/

Sorell, Tom and Heather Draper. 2014. “Robot Carers, Ethics, and Older People.” *Ethics and Information Technology* 16: 183–195. doi:10.1007/s10676-014-9344-7

Sparrow, Robert and Linda Sparrow, Linda. 2006. “In the Hands of Machines? The Future of Aged Care.” *Minds and Machines* 16 (2): 141–161.

Turkle, S., Taggart, W., Kidd, C. D., Dasté, O. 2006. “Relational Artifacts with Children and Elders: The Complexities of Cybercompanionship.” *Connection Science* 18 (4): 347–362.

Vandemeulebroucke, Tijs, Bernadette Dierckx de Casterlé, and Chris Gastmans. 2018. “The Use of Care Robots in Aged Care: A Systematic Review of Argument-Based Ethics Literature.” *Archives of Gerontology and Geriatrics* 74 (January): 15–25.

Wasserman, David. 2004. “Making Memory Lose Its Sting.” *Philosophy & Public Policy Quarterly* 24 (4): 12–18.

Wilson, R. S., Krueger, K. R., Arnold, S. E., et al. 2007. “Loneliness and risk of Alzheimer disease.” *Archives of General Psychiatry* 64 (2): 234–240. doi:10.1001/archpsyc.64.2.234

PART IV

TECHNOLOGY, METAPHYSICS, AND LANGUAGE

CHAPTER 15

THE TECHNOLOGICAL UNCANNY AS A PERMANENT DIMENSION OF SELFHOOD

CIANO AYDIN

1. Introduction

In previous decades, the view of technologies not being neutral has been defended from a range of perspectives (Ihde 1990; Latour 1992; Stiegler 1998; Feenberg 2002; Verbeek 2005). From these perspectives, technologies are seen not as merely neutral means developed by human beings to achieve certain goals that they have set for themselves. Rather technologies are attributed a power to co-shape both our world and our ideas, goals and values. They are shaping, according to some of these authors, even what it means to be human (Stiegler 1998; Verbeek 2005, 2011). Recognizing that technologies are normative and, hence, "norm" what we consider "successful or good self-formation" or an "enhanced self" has a far-reaching *existential* implication. Going beyond the inside-outside dualism and recognizing that what we consider our "inside" self is to a great extent shaped by our "outside" world implies that our "inside" is to a great extent also *for us* an "outside," which we cannot completely possess. Therefore, we cannot completely master and constrain our own process of self-formation. Or put differently: we do not completely possess the self that we attempt to form. It is not merely a "patient" that we can mold as we please.[1]

This sense of otherness within can be experienced, as I will discuss in this chapter, as uncanny. Our very selfhood seems to contain an otherness that cannot be simply externalized but is a constructive and structural part of what makes up who we are, which can elicit an eerie feeling. The question that I will address is how this otherness within that goes beyond the inside-outside distinction should be comprehended, whether there are more "voices," more types of "otherness" within the self—which is already suggested by the idea of a self that forms itself—and how these types of otherness relate to one another. The notion of the "uncanny" will be used to unravel these relations of alterity within, and to shed light on our existential condition in the light of a world saturated with technologies.

The concept of the "uncanny" has a history. In his seminal 1906 essay, *On the Psychology of the Uncanny* (*Zur Psychologie des Unheimlichen*), Ernst Jentsch takes as a starting point for his investigation of the uncanny the etymological meaning of the German word *unheimlich* (literally, "un-home-ly"), indicating that someone who experiences something uncanny is not quite "at home" or "at ease" in the situation concerned. The impression of the uncanniness of a thing or incident involves a "dark feeling of uncertainty," which is related to a "lack of orientation" (Jentsch 1906 [2008], 217, 224). Jentsch indicates that there is one exemplary experience that illustrates this uncanny feeling most clearly, namely the "doubt as to whether an apparently living being really is animate and, conversely, doubt as to whether a lifeless object may not in fact be animate" (Jentsch 1906 [2008], 221). For Jentsch, this is portrayed particularly in fiction, and more specifically, in storytelling. The lifelike doll Olympia, which features in E.T.A. Hoffmann's story "The Sandman" (*Der Sandmann*), is for Jentsch the prototypical example of an artifact that instigates a gloomy feeling of uncanniness (Jentsch 1906 [2008], 224).

The feeling of the uncanny that is brought about by automata was taken up in 1970 by the Japanese roboticist Masahiro Mori and designated as the "uncanny valley." Reviewing the different explanations of this "uncanny valley" will allow me to put forward an alternative interpretation of why encounters with humanlike automata elicit an eerie feeling. Hooking into how Jacques Lacan, via Sigmund Freud, takes up Jentsch's view of the uncanny, I will propose that uncanny feelings not only say something about our psychological responses to humanlike robots but also echo an ontological structure at the ground of human existence. Inspired by Lacan's notion of "extimacy," I will depict uncanniness as a fundamental dimension of our self-relation, as a permanent structure of subjectivity.

Lacan's notion of "extimacy" (Lacan 1997, 139; 2006, 224, 249) contributes to explaining why our capacity to form ourselves is restricted. This concept displays how the self is to a great extent a product of external influences and, therefore, cannot simply mold itself into whatever shape it pleases. However, Lacan's analysis primarily focuses on the symbolic order (language, laws, customs), not sufficiently taking into consideration the increasing impact of technologies on our self. Taking up Jean-Luc Nancy's concepts of "intrusion" (Nancy 2008, 161, 163, 167, 168, 169) and "being closed open" for technology (Nancy 2008, 168), I will illustrate how in our current era a technological order is ever more strongly shaping our selfhood. This technological order is "other" and "own" at the same time, which explains why technology can be experienced as uncanny.

Acknowledging that the technological uncanny is increasingly becoming a permanent structure of selfhood indicates that technology cannot simply be externalized and seen as an outside factor that can determine or liberate us, nor as something that can destroy or strengthen our autonomy. Both transhumanists who put their hopes on technologies that could enhance our physical and mental capacities and bioconservatives who warn us of the dangers of technologically tinkering with our biological and psychological make-up fail to sufficiently consider the implications of technology becoming "extimate." The proposed view calls for a more sophisticated account of how technology is shaping us, as well as how we would like to be shaped by it.

2. The Uncanny Valley

Ernst Jentsch (1906/2008, 223) already indicated that people confronted with clever automata are likely to grow more uneasy as the automata become more lifelike and refined. The more sophisticated the machine, the less confidence a spectator would have in drawing a line separating the animate from the inanimate. In his 1970 article entitled "The uncanny valley," the Japanese roboticist Mashihiro Mori depicted more precisely the relationship of familiarity and similarity in human likenesses and the positive or negative feelings that automata and other humanlike artifacts provoke. As a robot or other human duplicate becomes more human-like there is an increase in its acceptability, but as it approaches a nearly human state responses quickly turn to strong revulsion; as the robot's appearance continues to become less distinguishable from that of a human being, the emotional response becomes favorable once again (see Figure 15.1).

It should be noted that as the graph of the uncanny valley (Mori 2012, 99) flattens toward its peak, there is very little distance between the last instance where we still appreciate the robot's clever resemblance and the first disorienting moment that we feel repelled by its appearance. This "little distance" indicates that it is "minor differences" that instigate an uncanny feeling, an observation that can also be found in a different context in Freud, which I will take up later in this chapter.

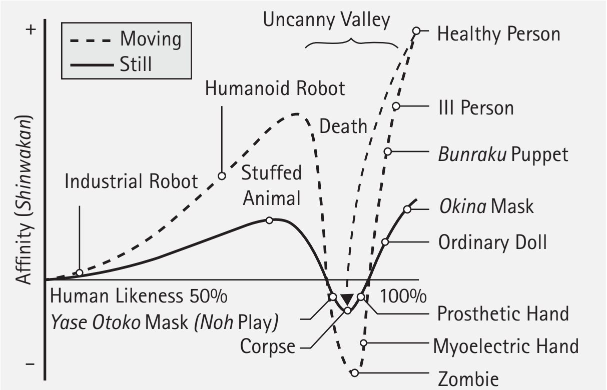

FIGURE 15.1: Mori's uncanny valley graph. The figure is taken from the 2012 translation of Mori's original paper: Mori, Masahiro. 2012. "The Uncanny Valley," Translated by Karl MacDorman and N. Kageki under authorization by M. Mori. *IEEE Robotics and Automation Magazine* 19: 99. Copyrights: Institute of Electrical and Electronics Engineers.

In designing humanoid robots, Mori advised to escape the uncanny valley by keeping a safe distance from complete human likeness (Mori 2012, 100). Instead of realistic eyes or hands that prompt uncanny feelings, designers, Mori recommends, should attempt to manufacture stylish devices that are sufficiently different from human faculties and, at the same time, could be easily and comfortably incorporated or related to (Mori 2012, 100). His advice has been taken up by engineers and filmmakers who, also for commercial reasons, try to avoid having their designs fall into the uncanny valley (Geller 2008). However, at the end of his paper Mori indicates—without further explanation—that his graph could also fulfill another function: "We should begin to build an accurate map of the uncanny valley so that through robotics research we can begin to understand what makes us human. This map is also necessary to create – using nonhuman designs – devices to which people can relate comfortably." (Mori 2012, 100). The order suggests that understanding what makes us human through an analysis of the uncanny valley is of even greater importance than creating "homey" robots. I will return to this later.

In later years, multiple studies sought to establish whether the uncanny valley is a real phenomenon and, if it is, to explain why it exists. Participants' ratings on familiarity or eeriness have been plotted against the human likeness of human replicas, using humanoid robots, androids and computer-generated characters; also morphing techniques have been employed to morph doll faces into human faces. Some of these studies show nonlinear relations that resemble the uncanny valley (MacDorman and Ishiguro 2006; Seyama and Nagayama 2007). A more recent study (Mathur and Reichling 2016) in which participants' ratings of 80 real-world android faces were observed and examined also detected a curve resembling the uncanny valley. However, other empirical studies did not detect nonlinear relations and, hence, did not confirm the uncanny valley hypothesis (Hanson 2005; MacDorman 2006; Bartneck, Kanda, Ishiguro, and Hagita 2007; Poliakoff, Beach, Best, Howard, and Gowen 2013). There are no rigorous controlled studies that unequivocally support the existence of the uncanny valley. However, there is support for its existence from a large number of more anecdotal studies and observations. Hence, whether the uncanniness of human-like artifacts is a function of their human like-ness remains debatable (Wang, Lilienfeld, and Rochat 2015, 394).

Multiple hypotheses have been proposed to explain the uncanny valley. Among these are a number of so-called perceptual theories. The *Pathogen Avoidance* hypothesis (MacDorman and Ishiguro 2006; MacDorman et al. 2009) was suggested by Mori himself, claiming that the uncanny valley must be related to "our instinct for self-preservation" (Mori 2012, 100). From this perspective, visual anomalies in human replicas, which are *perceived* as genetically very close to humans, elicit disgust because an evolved mechanism for pathogen avoidance detects these deficits as indicative of a heightened risk for transmissible diseases.

Alternatively, the *Mortality Salience* hypothesis (MacDorman and Ishiguro 2006) suggests that some humanlike robots remind human observers of their own inevitable mortality, thereby eliciting the uncanny feeling driven by a fear of death. Resembling dead people who move jerkily, humanoid automata would elicit the fear of being replaced by an android Doppelganger, being soulless machines, or losing bodily control (see also Ho, MacDorman, and Pramono 2008). Eerie feelings are explained in terms of defense systems that then are triggered to cope with that unpleasant prospect.

The *Evolutionary Aesthetics* hypothesis posits that humans are highly sensitive to visual aesthetics. This hypothesis suggests that selection pressures have shaped human preferences for certain physical appearances signaling fitness, fertility, and health. From this perspective, low attractiveness rather than lack of realism would explain the uncanniness of a human replica (Ferrey, Burleigh, and Fenske 2015; see also Hanson 2005).

In addition to perceptual theories, theories have been proposed that focus more on cognitive aspects to explain the uncanny phenomenon. The *Violation of Expectation* hypothesis was also suggested by Mori himself (2012), using the example of a prosthetic hand that appeared real at first sight but elicited eerie sensations as people realized that it was artificial. From this perspective, human replicas elicit an uncanny feeling because they create expectations but fail to match them (Mitchell et al. 2011). Here uncanniness is elicited not so much by how humanoids look but rather by how one thinks or assumes they will or should look. Saygin et al. (2012) suggested that a humanoid stuck inside the uncanny valley elicits repulsion because it is no longer judged by the standards of a robot doing a passable job of pretending to be human, but is instead judged by the standards of a human doing a terrible job of acting like a normal person.

The *Categorical Uncertainty* hypothesis goes back to Jentsch, who argued that uncanniness is associated with uncertainty and mistrust which generates disorientation. From this perspective, the uncanny phenomenon concerns the process whereby cognitive uncertainty emerges at any category boundary; negative affective responses are seen as a result of categorically ambiguous images, for example morphed images of a real, a stuffed, and a cartoon human face (Yamada, Kawabe, and Ihaya 2013).

The *Mind Perception* hypothesis addresses the question "On what bases do people perceive each other as humans?" From this perspective, the uncanny feeling is linked to violating the cognitive expectation that robots lack certain capacities that characterize humans, especially subjective experience, that is, the ability to feel and sense things (Gray, Gray, and Wegner 2007).

A theory that also focuses on robots coming too close to humans, instead of not close enough, is the *Threat to Distinctiveness* hypothesis, which suggests that humanlike robots, blurring category boundaries, undermine human uniqueness (Kaplan 2004; Ferrari, Paladino, and Jetten 2016). From this perspective, the fear of being replaced by a robot might not instigate fear of death but poses a threat to human identity, which elicits repulsion.

Wang, Lilienfeld, and Rochat (2015, 395f) have evaluated the validity of different perceptual theories and indicated that they suffer from limitations attributable to the methodologies used to test their hypotheses. Another problem they raise is the usage of morphed images or computer-generated characters, instead of existing human replicas, which forfeits, according to them, a certain degree of ecological validity. They have also evaluated cognitive theories that attempt to explain the uncanny feeling (2015, 397f) and pointed out that some theories of this kind neglect to explain what the cognitive expectations for humans and those for robots are, and why violating such expectations could elicit the uncanny feeling. They also note that cognitive theories fail to explain why attributing human feeling and sense experience to nonhuman and nonliving things, which belongs to a broader phenomenon known as anthropomorphism, does not seem to elicit negative effects in various other domains. In addition, Wang, Lilienfeld, and Rochat (2015, 398f) discuss conceptual difficulties in the translations and definitions of "uncanny" ("shinwakan" in Japanese) and "human likeness," and problems in measuring the dependent variable in the uncanny valley hypothesis. They suggest that unclear interpretations and conceptualization of the variables in the uncanny valley hypothesis may have contributed to inconsistent findings.

Wang, Lilienfeld, and Rochat (2015) stress the importance of studying the cognitive underpinnings of the uncanny phenomenon. They argue that many of the mentioned hypotheses provide plausible accounts of the uncanny phenomenon from different perspectives, while "they have neglected to verify the underlying assumption that observers would spontaneously perceive a human replica that closely resembles humans as a person" (2015, 401). Wang, Lilienfeld, and Rochat (2015) believe that this assumption is plausible, given the proclivity we have to anthropomorphize inanimate or nonhuman entities in literature, the arts, sciences, and in perception (Guthrie 1993).

Recognizing the cognitive process of anthropomorphism allows Wang, Lilienfeld, and Rochat (2015) to propose their own *Dehumanization* hypothesis. They argue that attributing humanlike characteristics to robots does not by itself explain the uncanny feeling; instead the uncanny feeling, they believe, must be understood as a response to a lack of humanness. An anthropomorphized human replica is not perceived to be a typical robot but is rather seen as a "robotlike" human. If the "robotlike" human then reveals its mechanistic nature, its humanness (above all the capacity for emotions and warmth) is questioned, which leads to dehumanization, thereby diminishing its likability and eliciting the uncanny feeling. This hypothesis is not necessarily in conflict with other hypotheses but interprets their findings from a different perspective: "The more human observers attribute humanlike characteristics to (i.e., anthropomorphize) a human replica, the more likely detecting its mechanistic features triggers the dehumanization process that would lead to the uncanny feeling" (2015, 402).

3. An Alternative Explanation of the Uncanny Valley, or the Importance of "Minor Differences"

The various hypotheses that I have listed above undoubtedly explain relevant aspects of the negative responses of certain humans to certain human-like robots. Moreover, Wang, Lilienfeld, and Rochat rightly show the plausibility of the assumption that in many studies observers tend to spontaneously perceive a human replica that closely resembles humans as a person. What is also noteworthy in relation to this assumption is that it is not the big but rather the *little* differences that evoke feelings of repulsion: observers spontaneously take humanlike robots as persons but are then repelled if they do not come close enough to humans, if small disparities reveal their lack of "humanness." The difference between having this "humanness" or not having it, seems to manifest itself in very subtle and elusive features: a small delay, an unexpected acceleration, an unfamiliar gesture. One moment the humanoid is human and the other he is not.

From a psychological perspective, the nonhuman, mechanistic traits of humanoids are primarily revealed in a lack of emotions and warmth, which, from this perspective, might be a sufficient explanation. However, from a more philosophical-existential perspective, the looming "little big" question is: "what makes this humanness"? What makes the ability to feel and sense "human"? Would we consider an android that perfectly possesses these capacities human? Or are these capacities mere surface markers of a deeper layer that designates a human? What is required to bridge the gap between a humanoid and a human? Often these questions lead to a kind of philosophical embarrassment: what makes us human seems to escape us. The psychological accounts of feelings of uncanniness seem to allow us to see something that may have otherwise remained hidden, something strange about our own identity and existence.

I am not the first to make the move from a psychological to a more existential-philosophical account of the uncanny. Katherine Withy (2015, 48) argues, building on Martin Heidegger, that the psychological accounts of the feeling that may accompany uncanniness refer to an "originary angst" that grounds falling (*Verfallen*), an "angst" expressing that the human cannot get a full hold of its own ground. From this perspective "humanness" is not characterized in terms of certain capacities that can be observed and measured but is, rather, rendered virtually *inaccessible*; as our mode of existence it is "too close to see." The feelings of uncanniness are interpreted as a fundamental mood that discloses a deeper ontological structure at the ground of human existence.

Yet instead of building on Heidegger, I would like to remain closer to the originators of the analysis of the uncanny. In his 1919 essay entitled *The "Uncanny,"* Sigmund Freud discusses and criticizes Jentsch's concept of the uncanny. He also draws on the work of Ernst Hoffmann and, like Jentsch, considers him the "unrivalled master of the uncanny in literature" (Freud 1981, 3686). In contrast to Jentsch, Freud did not regard almost-real objects *as such* as disturbing and dissonant, but rather believed that such feelings reveal deeper turmoil and psychopathology (Freud 1981, 3683). When Hoffmann's protagonist Nathaniel sees his object of love (the doll Olympia) partly dismantled with her eyes popped out of their sockets, thinks Freud, a repressed feeling resurfaced, namely the submerged fear of castration that survived from early childhood. Freud describes the uncanny as a "class of the frightening that leads back to what is known of old and long familiar" (Freud 1981, 3676), and, citing Schelling, as "*the name for everything that ought to have remained ... secret and hidden but has come to light* (Freud 1981, 3678)." For him Jentsch's conception of the uncanny is incomplete, since the recurrence of something repressed is required in order for a situation to be experienced as uncanny: without such resemblance, it can merely be frightening, which is different from uncanny. Freud stresses that this explains why the uncanny does not simply refer to something foreign but to an instance where something is foreign, yet disturbingly familiar at the same time. It is the "minor differences" that instigate a sense of uncanniness.

It is impossible and unnecessary to go here into questions regarding the validity of Freud's theory of repression. What I would like to take from Freud's approach is the idea that uncanniness revolves around the *tension* between unfamiliar and familiar, and hidden and revealed. Allowing us some freedom of interpretation and going outside of Freud's psychoanalytic framework, we might say that the humanlike robot elicits a feeling of uncanniness because it reveals something that *ought to have remained hidden*, namely the unfathomability of that which makes us human. The "minor difference" between the robot and the observer of the robot disorients not only because the robot is slightly different but also because what makes the observer different appears to be incomprehensible. From this perspective, the uncanny feeling is interpreted not only as a response to a lack of humanness in the robot, but also as a response to the viewer's own inability to fathom and appropriate this "humanness" that the viewer herself possesses.

In line with this view, I propose that the uncanny valley might say at least as much about us as it says about human-like robots. The robot might confront us with something uncanny in *us*. It is because a human-like robot resembles me without being completely identical ("minor differences") that I am confronted with my own unfoundedness, which is constitutively strange to me. I not only become aware of what makes me different from the robot but also of the impossibility for me to appropriate this difference. I do not suggest (nor exclude) that this explanation or interpretation could be validated by empirical research. Rather, I propose it as an explanatory, theoretical framework that could provide more insight into how technology is increasingly invasive and how our self has always been open for this technological intrusion. Following Lacan, who, via Freud, takes up the idea of the uncanny, will enable me to further elaborate the idea of these alterity relations within.

4. The Otherness of the Self as "Extimacy"

Freud uses the phrase "narcissism of minor differences" to show how it is the little differences "in people who are otherwise alike that form the basis of feelings of strangeness and hostility between them" (Freud 1981, 2355; see also 2553, 4506). Rudi Visker (2005, 433) explains that "narcissism" for Freud refers to an initially completely self-contained Ego that gradually opens up to reality. There is a movement from the inside to the outside: initially the Ego is a narcissistic entity exclusively focused on its libidinal drives, but then can gradually learn to redirect part of its energy and invest it in things outside itself. From this perspective, the self is originally a closed entity that can and should learn to gradually lose its protective shell and open up to the outside world and other people that, on first sight, seem strange and foreign. At this point, Visker (2005, 433) turns to Lacan. He notes that Lacan starts from the inverse hypothesis: the movement is not from the inside to the outside but from the outside to the inside. There is no closed original Ego, but rather the Ego is discovered and developed through the other.

Visker argues that connecting the notion of the "uncanny" to the concept of "narcissism of minor differences," which Freud himself did not explicitly do, and, via Lacan, reversing Freud's hypothesis, can foster insight into "alterity-relations within." This type of relation indicates that not only the otherness of other people needs to be recognized, as Emmanuel Levinas relentlessly stressed in *Totality and Infinity* (1961/1969) and other works, but also the otherness that somebody finds in herself.

In his famous essay *The Mirror Stage* (1949/2005) Lacan argues that the child discovers itself as a unified entity in and through something else, such as its own mirror image, the body of another child, and the responses of its parents. It would be inaccurate to say that the child recognizes itself in the Other, since it is only by virtue of that other and the discourses, goals, ideals and desires that others impart on it, that the child develops and discovers a self. From this perspective, identity is the result of identification, though without assuming that there is a subject prior to that process of identification (see also Julien 1990, 43–51).

Lacan's view of the relation between self and other is paradoxical and uncomfortable: the other is both the necessary condition for forming a self and at, at the same time, an obstacle that prevents the self from reaching the unity that it seeks. In Visker's words:

> identity will always bear the trace of an exteriority that it cannot fully interiorize. I am another (*je est un autre*) means: I cannot do without that other through whom I get an I. That other becomes someone that I cannot expel. In other words, my alienation is original, for it is implied in my self-constitution. There is no 'self-hood' without 'foreignhood.' The self is not something I possess, my 'self' is irremediably infected with an otherness that prevents me from being fully at one with myself.
>
> (Visker 2005, 433)

Instead of understanding the alterity within in terms of introducing another "in" the self, the self is revealed as something that is from the beginning contaminated with another. Lacan calls this otherness of the self *extimacy*: the "own"-ness of the self is both strange and familiar, both inside and outside, neither inside, nor outside. The self is always outside its center; the self is, one could say, referring to Helmuth Plessner's view that the human never completely coincides with herself, "ex-centric" (Plessner 1975).

Besides developmental psychological accounts of the self (such as the mirror stage), Lacan uses surrealistic and Escher-like figures to visualize the dizzying structure of extimacy, for example in the topology of the Möbius strip: the Möbius strip's half-twist results in an "odd" object (Lacan 2014, 120) because the single surface of the strip passes seamlessly from the "inside" to the "outside." Not only is it impossible to distinguish the inside surface from the outside one, but it is also impossible to tell left from right. It is disorienting and confusing: "You literally can't make heads or tails of it" (Robertson, 2015, 18). For Lacan, self-relations are characterized by this perplexing strain to distinguish "inside" from "outside."

From this Lacanian perspective, not only the other or otherness outside escapes definition—as Levinas (1961/1969) attempted to illustrate—but also the self that is confronted with that otherness. The self is not something I completely possess but is rather irremediably infected with an otherness within that prevents it from being fully at one with itself. The alienation is original, for it is implied in its self-formation. The self finds itself attached to something within, which is experienced as its selfhood, without being able to sufficiently understand and explain this attachment. It was *already there* before the self discovered itself as a self-reflecting agent. It cannot be fully objectified because it is always too close to the self.

This otherness within has for Lacan different dimensions. For one, the self's alterity within entails the influences of the external world that we have gradually incorporated. As we have indicated above, a child discovers and develops a unified self through embodying different external instances. The image that the self projects on itself through others is, according to Lacan, also an *imago*: a unified, stable and ideal totality ("that's you Helena, yes you are a wonderful girl, you are a princess, you're going to grow up to be beautiful and smart, just like your mommy"). The self attempts to realize this ideal image through identification, and, subsequently, enters a lifelong quest to correspond wholly with this Ideal-I, a quest that, Lacan stresses continuously, can never be completely fulfilled (Lacan 2005, 12, 15, 18). The *imago* also refers to the *imago Dei*, the image of God in which human beings were created and with which they should strive to conform but can never completely achieve. It is important to stress that the *imago* is not an emanation of the individual but the result of an encounter with larger Others and their desires, goals and ideals. Lacan sometimes designates this dimension, which also corresponds to a phase in the development of a child, as the "Imaginary Order" (Lacan 2005, 158, 161).

The images that others project on the self, by virtue of which it develops a sense of an unified Ego, also gradually enable the self to enter into what Lacan sometimes calls a Symbolic Order (and sometimes the "Big Other"): the pre-existing order of customs, institutions, laws, mores, norms, practices, rituals, rules, traditions, and so on of cultures and societies, which are entwined in various ways with language (Lacan 1997, 20, 81). The Imaginary and Symbolic do not coincide: the Imaginary is central to Lacan's account(s) of ego-formation and manifests itself in dyadic relations (such as in the self and its mirror image), whereas the Symbolic constitutes triadic relations by introducing, besides dyadic and intersubjective relation, a trans-subjective symbolic order that normatively regulates the relations between particular beings and society (Lacan, 1997, 81, 234). In short: the self is what it is in and through *mediations* of the endorsed image that others project on it, as well as through subjecting it to socio-linguistic arrangements and constellations.

There is besides the Imaginary and Symbolic Order also something else that constitutes the self, a dimension that Lacan designates with different names: the Thing (*La Chose*), the Real Other or the Real Order. In the *Mirror Stag*e (1949/2005) Lacan stated that in the image of the child reflected in the mirror there is one element, like the eyes of a creepy, living doll, that fails to integrate into a functional totality and necessarily appears fixed and immobile: the gaze. It has an uncanny way of detaching itself from me, said Lacan (Lacan 2014, 97; Robertson 2015, 25). It refuses to integrate into a functional totality. The reflection in the mirror serves to organize the child's movements and body parts in a unified whole. At the same time, this framing seems to leave behind a "residue" that escapes the subject's sense of complete mastery over her body (Lacan 2005, 3). Visker stresses that this drive within, unlike *angst* in Heidegger or shame in Levinas, is not something that is liberating or beneficial but an uncanny guest, a Thing that the self needs to be protected from. In all my attempts to control and "domesticate" it, I recognize that it escapes me, might cross the borders that I and society have set, disorient me, and potentially might destroy me (Lacan 1997, 43–56).

However, the Thing is also not sheer negativity, as Lacan's depictions of the Thing might suggest. It is something that does not fit and cannot fit into an encompassing frame of meaning. By virtue of this aspect, the self can never be completely captured and domesticated by the ideal images that others project on it, nor by the symbolic order in which it is immersed. It is this dimension that gives the self particularity and singularity.

This characterization of the self renders Lacan's psychological anthropology completely at variance with Anglo-American ego psychology and the Enlightenment spirit, which seek to strengthen people's ego and liberate them from restrictions. Despite having consciousness, the self is not a locus of autonomous agency, it is not the seat of a free "I" determining its own fate. The self is thoroughly *compromised*. The other (in its three manifestations as Imaginary, Symbolic, and Real) is both the necessary condition for forming a self and an obstacle that prevents the self from reaching the unity, autonomy and singularity that it seeks, not only because it cannot meet certain demands of others or because it has been shaped by a world that it was thrown into (to borrow a Heideggerian term) but also because it can never fully appropriate what it desires. Without the other it is impossible to discover and develop subjectivity or selfhood and, at the same time, the other prevents it from becoming an autonomous being, unaffected by its traces, inscriptions and whims; or put yet differently, in all my attempts to become an independent and unique self, I remain to a great extent a repository for the projected desires and fantasies of larger others and a plaything of the idiosyncratic and disruptive vagaries of an unruly force within.

This makes, as indicated earlier, the "otherness" in the self more disturbing, since the self is unable to externalize it, detach itself from it and localize it, which explains its uncanniness. Since the self becomes what it is by virtue of its encounters with manifestations of this other, it remains always a stranger or other for itself. The self is never completely "at home." Its "own"-ness is, as we have seen, both strange and familiar, which explains why uncanniness is a permanent dimension of its subjectivity.

It is crucial to understand this other within from a radical anti-essentialist view that goes beyond inside-outside dualisms, since the other that is beyond our control is, at the same time, responsible for forming our self; our self-relation is inherently an alterity relation. The self is, contra Freud, not something that has to learn to open itself for others, but rather has to find a way to live and not to be crushed by that other that, from the beginning, is already inside: the stranger outside me can make me aware of and awakens the otherness inside me, which can fill me with incongruity, confusion and, sometimes, rage (see also Visker 2005, 435).

5. Being "Closed Open" for Technology

Lacan illustrates how otherness structurally constitutes the self. In his depiction of how society shapes the self he predominantly focuses on the world of language, laws and customs. However, in our present culture, we are witnessing, besides or in addition to a Symbolic Order, the ever-stronger ubiquity of a Technological Order. Today virtually all facets of our lives are saturated with technology. It must be said that the material world is not absent in Lacan's account. It is notable that his prime example of Otherness involves an artifact, namely the "*mirror* image." In fact, in the 1949 text, Lacan seems to think of artifacts as equally relevant props as humans (parents, peers, etc.) within the context of subjectivation but in subsequent reinterpretations of the mirror stage during the 1960s, he increasingly highlights the supporting role of fellow human beings, caregivers' narratives, and socio-linguistic factors. If Lacan would have lived and written in our era, where technologies are becoming more intimate and intrusive than ever before, he probably would have emphasized more the role of technologies such as screens, tablets, mobile phones, social networking services, brain imaging and other medical technologies, and algorithms and other digital grammars. In order to explain how we find ourselves in an "extimate" relation not only with a symbolic but also with a technological order, and how this relation is increasingly shaping our selfhood, I will complement Lacan's notion of "extimacy" with Nancy's view of being "closed open" for technologies. This technological order that is "other" and "own" at the same time, might further explain in what sense technology is experienced as uncanny in our current era.

In 1990 the French philosopher Jean-Luc Nancy got severely ill and needed to undergo a heart transplant. In an autobiographical essay entitled *L'intrus* (*The Intruder*) he documented this experience. Nancy notes that his heart has always seamlessly kept him alive, supplying oxygen and nutrients to the tissues in his body. Before his illness, his heart was, as Nancy describes, the most private and intimate part of himself and, at the same time, not more than a piece of meat, invisible and without meaning. After he got ill, his relation to his heart radically altered: in order to survive, he had to get rid of it. Nancy says: "My heart became my stranger" (2008, 163). Nancy was still his heart but, at the same time, his heart became something foreign. Instead of an ally, suddenly his heart became a dangerous enemy. His heart became an intruder, not one that enters from outside but one that enters from *inside* (Nancy, 2008, 162f.). We see here that the idea of an "intruder from inside" renders the apparently clear-cut distinction between "inside" and "outside" opaque.

Besides his sick heart, Nancy describes many other forms of strangeness that he experienced. The donor heart that he got was seen as a stranger. As Nancy states: "my heart can be a black woman's heart" (Nancy 2008, 166). Also his own immune system—normally his most important protector and ally—became a threat, since it needed to be suppressed in order to accept the donor heart. Furthermore, his age became a stranger, since the donor heart could be twenty years younger than he is (Nancy 2008, 169). And this was not the end of Nancy's strange encounters: after his heart transplant, Nancy got sick again and developed cancer; now the cancer cells, which prior to his illness were not identified as different, became a dangerous stranger.

The long list of strange entities that he came across in his body led to Nancy's observation that not only parts of his body but also his body *as such* is a stranger to him. Moreover, while reading Nancy's essay it gradually becomes clear that its main theme is not his heart transplant, nor his cancer cells, nor his illness. His line of thought culminates in a reflection about how the "intruders" from within and without reframe his view of his "self." He writes:

> I am the illness and the medicine, I am the cancerous cell and the grafted organ, I am these immuno-depressive agents and their palliatives, I am these ends of steel wire that brace my sternum and this injection site permanently sewn under my clavicle, altogether as if, already and besides, I were these screws in my thigh and this plate inside my groin.
>
> (Nancy 2008, 170).

In addition, Nancy's focus shifts from observations on his body and the way he relates to it to a reflection on the technologies that are inserted in his body, the technological manipulation of his body, and how his relation to these technologies sheds a different light on his body and self.

In relation to the notion of the "intruder," Nancy employs the idea of the self being "closed open" (Nancy 2008, 168) which together signify how the technologies that are used to treat and keep his body alive are ever more interwoven with his very self: " 'I' always find itself tightly squeezed in a wedge of technical possibilities" (Nancy 2008, 162). The idea of being "closed open" indicates that the technologies used to treat Nancy should not be seen as strangers from an outside realm that infringe the self; rather the self is exposed as always having been part of that "outside." As Nancy explains: "What a strange me! Not because they [the surgeon, the technologies] opened me up, gaping, to change the heart. But because this gaping cannot be sealed back up. (. . .) I am closed open" (Nancy 2008, 167f.).

Nancy stresses that current technologies highlight the alterity in selfhood, though, at the same time, he makes clear that they did not cause or generate it: "never has the strangeness of my own identity, which for me has always been nonetheless so vivid, touched me with such acuity" (2008, 168; see also Slatman 2007). Nancy attempts to illustrate, very much in line with Lacan, that alterity is a constant dimension of our self and self-experience. It can also be experienced if the body is not ill. The heart transplant and other technological intruders make this experience only more acute, but have not generated this being "closed open." We have always been strangers to ourselves. In Nancy's words:

> The intruder is nothing but myself and man himself. None other than the same, never done with being altered, at once sharpened and exhausted, denuded and overequipped, an intruder in the world as well as in himself, a disturbing thrust of the strange, the *conatus* of an on-growing infinity.
>
> (Nancy 2008, 170)

The self has always been outside itself and, hence, can never be completely *closed* in order to entirely possess itself.

It is clear that for Nancy (and for Lacan) the self has always been "closed open," but that does not imply that with the advent of new technologies there is nothing new under the sun. New and emerging technologies have expanded the possibilities to "intrude" in the "closed open self," which is also confirmed by Nancy: "I am turning into something like a science-fiction android, or else, as my youngest son said to me one day, one of the living-dead." (Nancy 2008, 170) Besides tradition, education and culture, now technology has become a dominant force in self-formation processes, as Nancy very intimately has experienced. The human has always been "closed open" but now she can immediately intervene in her own bodily constitution. The potential to be "closed open" has always existed, but technologies today take increasing advantage of this potentiality:

> Man becomes what he is: the most terrifying and the most troubling technician, as Sophocles called him twenty-five centuries ago, who denatures and remakes nature, who recreates creation, who brings it out of nothing and, perhaps, leads it back to nothing. One capable of origin and end.
>
> (Nancy 2008, 170).

Lacan's idea of "extimacy" highlights that the self is not a closed "inside" that then learns to open up to the outside world, but that the self is rather discovered and developed in and through a pre-existing symbolic order, an order that, on the one hand, is constitutive for its subjectivity and agency and, on the other hand, is an obstacle that prevents it from reaching the autonomous unity that it desires. What Nancy's elaboration of his experience of being "closed open" for technologies adds to this framework, is that the self, as an epistemic object, is ever more deeply immersed in a pre-existing realm of biomedical knowledge and technology. This technological realm is increasingly shaping the self. The technology that potentially always can intrude in the self also affects and transforms how the self experiences itself, and in which direction the self is formed. For Nancy, technology does not extend the mind or the self but the self has always been open and exposed and now technology is excessively *confiscating* it (see also Aydin 2015): "the subject's truth is its exteriority and its excessiveness: its infinite exposition. The intruder [in this case technology] exposes me to excess. It extrudes me, exports me, expropriates me" (Nancy 2008, 170).

Complementing Lacan's notion of "extimacy" with Nancy's view of being "closed open" for technologies makes it possible to reinterpret the other within in terms of technology within. The technology within is not completely strange or foreign, since it is a constitutive part of our subjectivity and selfhood. At the same time, technology prevents one from becoming an autonomous and singular being, unaffected by its engravings. Technology is strange and familiar, at the same time. That "at the same time" explains why it can be experienced as uncanny.

6. Alterity in Selfhood and the Question of Technological Self-Formation

The idea of the uncanny has been used to designate an alterity within that cannot be simply explained in terms of something external that challenges or influences our internal convictions, preferences, values, or goals. From Freud I have taken the view that the uncanny cannot be simply be opposed to the canny: *heimlich* and *unheimlich* are not simply opposites, since *unheimlich* signifies the concealed and the hidden and, at the same time, the familiar and domestic. The uncanny within is strange and familiar, at the same time.

Lacan's notion of "extimacy" has been employed to further illustrate how "ownness" does not exclude but rather includes "otherness." This notion expresses, on the one hand, that even our most personal goals, aspirations and ideals that we attempt to realize in order to become an ideal-I are derived from significant others in our lives. The sense of being a unified Ego is derived from images that others project on me. My desires are ultimately desires of others, such as my parents, educators, role models, superstars, Party, God, Nature, and Science. In confrontation with significant others we gradually enter a symbolic order that enable us to become part of a community and define ourselves from a third person perspective as subjects with certain roles, duties and responsibilities.

On the other hand, Lacan stresses that there is also some-Thing in us that prevents being completely absorbed by societal aspirations, values and ideals, including ethical, political, and (we can add) *technological* rules, regulations and grammars. Although by virtue of this drive humans are singular beings, Lacan points out that this dimension should not be romanticized; in its purest form it is an unfathomable and disorienting abyss of withdrawn-yet-proximate alterity. In order to regulate its drives and impose a form to them, the self needs help in the form of a symbolic and technological frame or narrative. Lacan stresses the importance of the Symbolic and Imaginary and their protective, orienting and stabilizing workings. At the same time, he points out that the process of subjectivation and socialization always hold the chance of excessively repressing and fixing the self through a particular, "sheltered" system that ultimately becomes a straitjacket and prevents developing a singular identity.

Through a reading of Nancy's *Intruder* I have tried to complement Lacan's social order of language, laws and regulations with a technological order that is increasingly shaping the very nature of our selfhood. In our current era technologies and technological systems can be added to the Lacanian Imaginary Other that projects its desires on us, and the "Big Other" that regulates our conduct: an iPhone is not only a handy device for making calls, texting and surfing the web, but promises us to upgrade our identity and lifestyle. Brain imaging technologies are increasingly used not only to diagnose diseases and lesions but also to correlate brain activation with psychological states and traits, up to a level that, some predict, will enable us to correct the mental states that someone ascribes to herself or even establish whether someone really possesses free will (Aydin 2018). Upcoming persuasive technologies will influence our wishes and desires more seamlessly, making it even harder to recognize them as being projected to us (Frischmann and Selinger 2018).

The idea of a socio-technological order influencing and regulating our conduct and interactions, as well as generating social stability, is not a completely novel view. Philosophers like Hegel and Gehlen have argued that institutions and institutionally conveyed mental habits have the formal and informal function to unburden and give coherence and continuity, to compensate for the human's lack of instinctual determination.

However, Lacan and Nancy illustrate that this order is a constitutive dimension of the self, and cannot be simply externalized and objectified. Instrumentalist and determinist approaches to technology, as well as techno-optimist and techno-pessimist approaches (including transhumanist and bioconservative approaches) often overlook that technology cannot be simply situated outside humans and their condition. The "technological other" limits our capacity to form ourselves not because it constrains an original capacity to make autonomous decisions, but because this "technological other" has engraved—and is ever more deeply engraving—its structures in our very origin. Technology increasingly enables us to form ourselves into stable and socially dependable beings and, at the same time, prevents us from reaching the autonomy and uniqueness that we seek, which could account for the uncanniness that some technology seems to elicit.

Reflection on the uncanny feeling triggered by a humanlike robot prompts the question not only of what makes robots different from humans, but also what makes humans different from robots: the lack of humanness that would elicit the uncanny feeling instigates the question of what makes up this "minor difference." In confrontation with the humanlike robot I not only become aware of what makes me different from it, but also of the impossibility for me to appropriate that difference. The elaboration of the "extimate structure" of the self has led to the finding that the self is formed in the image that the "outside world" projects on it. Since the "outside world" is increasingly a world of technology, the self, being "closed open," is increasingly being shaped in the *image* of technology; technology is increasingly becoming the "Big Other," the dominant "intruder within." Thus, perhaps in the confrontation with the strange and, at the same time, familiar robot, the self not only uncannily senses the human in the robot but also *the robot in the human*. From this view, it is inaccurate and inadequate to frame the self as something that could or should close itself off from, or alternatively learn to open itself for, technology. In line with what Lacan and Nancy say about the other within, I propose instead that the self should find a way not to be restricted and crushed by, but rather live in a *deliberate* way with, the technology which from the beginning is already inside.

However, there is a complicating factor which I have ignored so far. For Lacan the social order, on the one hand, protects the self from arbitrariness and excess and, on the other hand, always comes with the chance that protection keels over to repression. What Lacan does not seem to have envisaged, besides the view that that the symbolic other could be toppled by a technological other, is that this technological other, instead of securing *order*, could also become a source of disruption itself. For example, transhumanists and other techno-optimists who propose enhancing human capacities by means of existing and emerging technologies often do not take into account that these technologies are influencing, challenging and disrupting our very standards for establishing what *are* "enhanced capacities." They wrongly assume that it is possible to refer to univocal standards for measuring what is "disabled" and "normal," as well as what is an enhanced self or "successful or good self-formation" (Aydin, 2017). In addition, the authority of traditional "Big Others" is more easily questioned and challenged in our current global society, which harbors different and sometimes opposing views, values and ideals, different and opposing views that are accessible to ever greater parts of the world population through the Internet and other media. Not only do we need to deal with the human being as a "monster and an abyss," that is, a being that escapes every possible uniform categorization and, therefore, continuously is able to challenge and disrupt our standards, we now also seem to witness a "technological other" becoming an additional disruptive force. The technological other is becoming an additional disorienting dimension that could further intensify the uncanny within.

In the wake of univocal standards being challenged and undermined from different others without and within, the question of how to form ourselves becomes ever more acute. How is it against this background still possible to sustain the ideal of "good self-formation"? How can one develop both a coherent and a singular self in the light of our intrinsic technological condition? Recognizing that technology "conditions" our humanness, could we also consciously employ it to "condition" our humanness in a certain desired direction and form ourselves in a "good" way? I believe that the notion of "sublimation" might prove itself fruitful in this respect, but that is a topic for a later study.

Note

1. In this chapter I use the concepts of "self" and "identity" interchangeably, and not in the more technical-analytic fashion that we can find in debates about personal identity, agency, self-identity, etc. Their meanings should be derived from the elaborated theories and views.

References

Aydin, Ciano. 2015. "The Artifactual Mind: Overcoming the "Inside-Outside" Dualism in the Extended Mind Thesis and Recognizing the Technological Dimension of Cognition." *Phenomenology and the Cognitive Sciences* 14, no 1: 73–94.

Aydin, Ciano. 2017. "The Posthuman as Hollow Idol: A Nietzschean Critique of Human Enhancement." *Journal of Medicine and Philosophy* 42, no 3: 304–327.

Aydin, Ciano. 2018. "From Camera Obscura to fMRI: How Brain Imaging Technologies Mediate Free Will." In *Postphenomenological Methodologies: New Ways in Mediating Techno-Human Relationships*, edited by Jesper Aagaard, Jan Kyrre Berg Friis, Jessica Sorenson, Oliver Tafdrup, and Cathrine Hasse,103–122. Lanham/Boulder/New York/London: Lexington Books.

Bartneck, Christoph, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2007. "Is the Uncanny Valley an Uncanny Cliff?" In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication*, 368–373. New York: IEEE Press.

Feenberg, Andrew. 2002. *Transforming Technology*. New York: Oxford University Press.

Ferrari, Francesco, Maria Paola Paladino, and Joland Jetten. 2016. "Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness." *International Journal of Social Robotics* 8, no 2: 287–302.

Ferrey, Anne, Tyler J. Burleigh, and Mark J. Fenske. 2015. "Stimulus-Category Competition, Inhibition, and Affective Devaluation: A Novel Account of the Uncanny Valley." *Frontiers in Psychology* 6, no. 249. doi:10.3389/fpsyg.2015.00249

Freud, Sigmund. 1981. *The Standard Edition of the Complete Works of Sigmund Freud (CPW)*, edited by J. Strachey. London: Hogarth Press.

Frischmann, Brett, and Evan Selinger. 2018. *Re-engineering Humanity*. Cambridge: Cambridge University Press.

Geller, Tom. 2008. "Overcoming the Uncanny Valley." *IEEE Computer Graphics and Applications* 28: 11–17.

Gray, Heather M., Kurt Gray, and Daniel. M. Wegner. 2007. Dimensions of Mind Perception. *Science*, *315* (5812): 619. doi:10.1126/science.1134475

Guthrie, Stewart. 1993. *Faces in the Clouds*. New York, NY: Oxford University Press.

Hanson, David. "Expanding the Aesthetic Possibilities for Humanoid Robots." In *IEEE-RAS International Conference on Humanoid Robots*, 2005. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.2518andrep=rep1andtype=pdf

Ho, Chin-Chang, Karl F. MacDorman, and Z. A Dwi Pramono. "Human Emotion and the Uncanny Valley: A GLM, MDS, and Isomap Analysis of Robot Video Ratings." In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, 169–176, 2008. doi:10.1145/1349822.1349845

Ihde, Don. *Technology and the Lifeworld*. Bloomington: Indiana University Press, 1990.

Jentsch, Ernst. "On the Psychology of the Uncanny." In *Uncanny Modernity: Cultural Theories, Modern Anxieties*, edited by J. Collins and J. Jervis, 216–228. London: Palgrave Macmillan 1906/2008.

Julien, Philippe. 1990. *Le retour* à *Freud de Jacques Lacan*. Paris: E.P.E.L.

Kaplan, Frederic. 2004. "Who Is Afraid of the Humanoid? Investigating Cultural Differences in the Acceptance of Robots." *International Journal of Humanoid Robotics*, 1, no. 3: 465–480.

Lacan. Jacques. 1997. *The Ethics of Psychoanalysis 1959–1960. The Seminar of Jacques Lacan Book VII*. Edited by Jacques-Alain Miller, translated by Dennis Porter. New York: Norton.

Lacan, Jacques. 2005. *Écrits: A Selection*. Translated by Alan Sheridan. London: Routledge.

Lacan, Jacques. 2006 (1968–1969). *Le séminaire XVI: D'un Autre* à *l'autre*. Paris: Éditions du Seuil.

Lacan, Jacques. 2014. *Anxiety*. *The Seminar of Jacques Lacan Book X*. Edited by Jacques-Alain Miller, translated by A. R. Price. Cambridge: Polity Press.

Latour, Bruno. 1992. "Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts." In *Shaping Technology/Building Society*, edited by Wiebe E. Bijker and John Law, 225–258. Cambridge, MA: MIT Press.

Levinas, Emmanuel. 1961/1969. *Totality and Infinity*, *an essay on exteriority*, translated by Alphonso Lingis. Pittsburgh: Duquesne University Press.

MacDorman, Karl. 2006. "Subjective Ratings of Robot Video Clips for Human Likeness, Familiarity, and Eeriness: An Exploration of the Uncanny Valley." In *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, Vancouver. http://www.coli.uni-saarland.de/courses/agentinteraction/contents/papers/MacDorman06short.pdf

MacDorman, Karl F., and Hiroshi Ishiguro. 2006. "The Uncanny Advantage of Using Androids in Cognitive and Social Science Research." *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* 7, 297–337.

MacDorman, Karl F., Robert Green, Chin Chang Ho, and Clinton T. Koch. 2009. "Too Real for Comfort? Uncanny Responses to Computer Generated Faces." *Computers in Human Behavior*, 25: 695–710.

Mathur, Maya B., and David B. Reichling. 2016. "Navigating a Social World with Robot Partners: A Quantitative Cartography of the Uncanny Valley." *Cognition*, 146: 22–32.

Mitchell, Wade J., Kevin A. Szerszen, Sr., Amy Shirong Lu, Paul W. Schermerhorn, Matthias Scheutz, and Karl F. Macdorman. 2011. "A Mismatch in the Human Realism of Face and Voice Produces an Uncanny Valley." *i-Perception* 2: 10–12.

Mori, Masahiro. 2012. "The Uncanny Valley," Translated by Karl MacDorman and N. Kageki under authorization by M. Mori. *IEEE Robotics and Automation Magazine* 19: 98–100.

Nancy, Jean-Luc. 2008. "The Intruder." In *Corpus*, translated by R. Rand. New York: Fordham University Press.

Plessner, Helmuth. 1975. Die Stufen des Organischen und der Mensch: Einleitung in die philosophische Anthropologie. Berlin-New York: De Gruyter.

Poliakoff, Ellen, Natalie Beach, Rebecca Best, Toby Howard, and Emma Gowen. 2013. "Can Looking at a Hand Make your Skin Crawl? Peering into the Uncanny Valley for Hands." *Perception* 42: 998–1000.

Robertson, Brian. 2015. *Lacanian Antiphilosophy and the Problem of Anxiety: An Uncanny Little Object*. New York: Palgrave Macmillan.

Seyama, Jun'ichiro and Ruth S. Nagayama. 2007. "The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces." *Presence* 16: 337–351.

Saygin, Ayse Pinar, Thierry Chaminade, Hiroshi Ishiguro, Jon Driver, and Chris Frith. 2012. "The Thing that Should Not Be: Predictive Coding and the Uncanny Valley in Perceiving Human and Humanoid Robot Actions." *Social Cognitive and Affective Neuroscience* 7: 413–422.

Slatman, Jenny. 2007. "Grenzen aan het Vreemde." *Wijsgerig Perspectief* 47: 6–16.

Stiegler, Bernard. 1998. *Technics and Time, 1. The Fault of Epimetheus*. Stanford, CA: Stanford University Press.

Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park: Pennsylvania State University Press.

Verbeek, Peter-Paul. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press.

Visker, Rudi. 2005. "The Strange(r) within Me." *Ethical Perspectives* 12, no 4: 425–441.

Wang, Shensheng, Scott O. Lilienfeld, and Philippe Rochat. 2015. "The Uncanny Valley: Existence and Explanations." *Review of General Psychology* 19: 393–407.

Withy, Katherine. 2015. *Heidegger on Being Uncanny*. Cambridge: Harvard University Press.

Yamada, Yuki, Takahiro Kawabe, and Keiko Ihaya. 2013. "Categorization Difficulty Is Associated with Negative Evaluation in the 'Uncanny Valley' Phenomenon." *Japanese Psychological Research* 55: 20–32.

CHAPTER 16

TECHNOLOGY AND THE ONTOLOGY OF THE VIRTUAL

MASSIMO DURANTE

> . . . *les personnages, les objets, les images, et d'une manière générale tout ce qui constitue la* réalité virtuelle *du théâtre*
> Antonin Artaud, Le théâtre et son double, 1938, 49.

1. Introduction

What is real? What is virtual? In posing these questions, we risk sliding down a rabbit hole. Is the young couple sitting across from me real? How can one possibly doubt it? What about Romeo and Juliet? Are they real, too? Certainly no one would attribute to Romeo and Juliet the same degree of being as they would to this couple, would they? Although we are much more familiar with Romeo and Juliet and have experienced strong emotions on account of their tragic tale, I believe that most of us would nonetheless consider the couple in front of me to be real in a way that Romeo and Juliet are not. And yet the thoughts and feelings we have about Romeo and Juliet are indeed real, are they not? Hence, can something virtual produce real consequences? Is there a definite and unequivocal dividing line between what is real and what is virtual? And finally, is there a difference between "the virtual" and what we refer to as "virtual reality"? In other words, is there a difference between the virtual in analog reality and the virtual in digital reality?

As human beings, we have always dealt with the virtual. Language itself is perhaps the most elemental and powerful instrument for producing virtual reality, in its ability to evoke intangible and non-present entities. As Judea Pearl and Dana McKenzie recently recalled, referring to Yuval Harari, the main evolutionary asset of human beings is their ability to imagine things that do not exist:

> I totally agree with Yuval Harari that the depiction of imaginary creatures was a manifestation of a new ability, which he calls the Cognitive Revolution. His prototypical example is the Lion Man sculpture [. . .]. As a manifestation of our newfound ability to imagine things that never existed, the Lion Man is the precursor of every philosophical theory, scientific discovery, and technological innovation, from microscopes to airplanes to computers.
>
> (Pearl and McKenzie 2018, 34–35)

> We cooperate effectively with strangers because we believe in things like gods, nations, money and human rights. Yet none of these things exists outside the stories that people invent and tell one another. There are no gods in the universe, no nations, no money and no human rights—except in the common imagination of human beings.
>
> (Harari 2015, 28)

Plato's cave is a celebration of virtual reality.[1] The second and third chapters of Hobbes' *Leviathan*—the text on which political modernity is built—are respectively devoted to imagination and the train of imagination. The Kantian regulative ideal is a critical instance of reality by virtue of a dimension that is, precisely, only virtual. In short, the virtual existed long before and regardless of any technologically digitized virtual reality.

Nevertheless, the evolution of technology—and digitization in particular—has altered this scenario. Digitization forces us to recognize that the virtual has acquired its own independent form of reality, namely, virtual reality. The virtual has thus become part of the real and vice versa, as Luciano Floridi has well observed:

> What matters is not so much moving bits instead of atoms—this is an outdated, communication-based interpretation of information society that owes too much to mass-media sociology—as the far more radical fact that our understanding and conceptualization of the essence and fabric of reality is changing. Indeed, we have begun to accept the virtual as partly real and the real as partly virtual.
>
> (Floridi 2014a, 218)

With its massive development of software systems, artificial agents, and the algorithmic outsourcing of actions and decisions, digitization also leads us to an even more basic realization: different epistemic agents—either human or artificial—encounter real and virtual reality at different levels of abstraction, to take up a phrase coined by Floridi (2008). For an artificial agent, which processes data, that which we refer to as virtual reality is the only reality. It is we alone for whom such a reality is virtual. No matter how we define virtual reality, definitions are meaningful only to human beings. In other terms, "As technology changes everything, we have a chance to discover that by pushing tech as far as possible we can rediscover something in ourselves that transcends technology" (Lanier 2017, 55).

Seeking endless definitions of the virtual, may thus be futile or perhaps even counterproductive. What is important to grasp is that, through the virtual, we explore alternative modes of experiencing reality or, better said, we probe the limits of our experience. Furthermore, digitization brings with it a key novelty: the virtual is no longer an epistemic construction that exists exclusively for the benefit of human beings. Virtual reality is not merely populated by objects, states, events, and actions that serve human purposes. Machines adapt the representation of the world to their own functioning and build up a virtual reality that provides the most suitable environment for their operations.

In simple terms, this chapter aims to show that an ontology of the virtual that is independent from technology no longer exists. From an epistemic standpoint, different agents encounter the different modes of reality they are able to construct. The main challenge to dealing with the issues at stake stems from the conventional, materialist ontological approach taken to them, or in other words, from attempting to answer the set of questions posed at the start of this chapter as if they were based on some materially identifiable essence or substance. The fallacy lies in considering such questions as "what-questions" as opposed to "what-for questions," that is, questions that investigate and revolve around purposes, roles, and functions. The next section discusses a more helpful ontological approach to take.

2. What and What-for Questions

When I was asked to write this chapter on the technology and ontology of the virtual, I immediately recalled an important lesson I learned long ago from Jean-François Courtine (1990): Ontology is a modern word. It has little to do with the philosophy of ancient Greece. Modern ontology is distinct from traditional metaphysics. The first occurrences of the term date from the early seventeenth century, first in Jacob Lorhard (1606) and then in Rudolfh Göckel (1613). Its conceptual development owes much to Johannes Clauberg (1647) and its first systematic theorization can be found in Christian Wolff, as Etienne Gilson (1952, 112) correctly pointed out:

> This text may be held, in the present state of historical knowledge, for the birth certificate of ontology as a science conceived after the pattern of theology, yet radically distinct from it, since being qua being is held there as indifferent to all its conceivable determinations. There is, Clauberg says, a certain science which envisages being inasmuch as it is being, that is, inasmuch as it is understood to have a certain common nature or degree of being, a degree which is to be found in both corporeal and incorporeal beings, in God and in creatures, in each and every singular being according to

> its own mode. Leibniz will later praise Clauberg for such an undertaking, but he will regret that it had not been a more successful one. The very word 'ontology' occurs at least once in an undated fragment of Leibniz, and one can expect accidentally to meet it later in various places, but it is not until 1729 that it finally comes into its own with the Ontologia of Christian Wolff.

Although it is not the task of the present chapter to provide a full discussion of the notion of ontology itself, there is one important aspect that bears mentioning. Ontology asserts itself in modernity with the precise aim of including in the investigation of what exists—reality or existence—that which is merely intelligible ("intelligible as intelligible" according to Clemens Timpler's formula, 1604): namely, that which is immaterial, incorporeal or nonphysical, alongside that which is material, corporeal and physical.

> Following Timpler, Lorhard defined ontology as 'the knowledge of the intelligible by which is intelligible.' His ontology is hence a description of the world of intelligibles, i.e., the items, concepts, or objects that are understandable or conceivable from a human perspective. The emphasis on the intelligibility of the world is essential in Timpler's and Lorhard's ontology. When Lorhard followed Timpler's lead and adopted this new proposal about the subject matter of metaphysics, or ontology, he agreed with the idea that we in formulating ontology are concentrating on the knowledge by means of which we can conceive or understand the world. In this way ontology is seen as a description of the very foundation of scientific activity as such.
>
> (Øhrstrøm, Schärfe, and Uckelman 2008, 76)

In short, ontology has always encompassed a constant and profound meditation on the virtual. In this sense, the virtual has been understood in modernity as the benchmark—or epistemological limit—against which the real is measured. Modernity itself has essentially been characterized in epistemological terms. It is epistemology that decides how and which realities can be known: ontology has been consistently defined by the level of knowledge available at each stage of the modern age, and the virtual has accordingly tested the limits of that which can be known through sensible experience.

At the peak of its development, the virtual has indeed been included in the predicate of existence, and it has long been possible to speak of a virtual reality, that is to say, the fact that the virtual exists in its own right. For this reason—and for another reason shortly to be explained—it would be rather impractical to devote ourselves to defining what is virtual. Such a definition would require us to be able to accurately distinguish the real from the virtual (i.e., the material from the immaterial, the corporeal from the incorporeal, the physical from the nonphysical), taking into account the entire evolution of the notion of reality across modernity. That would exceed the scope of this text.

There is an additional and even more pressing argument for choosing not to frame the ontology of the virtual in terms of a what-question. It lies in the changing of the criterion of existence (which is a particular challenge not only to traditional metaphysics but also to modern ontology), as Luciano Floridi has emphasized. In the digital age, the criterion of existence itself has dramatically changed as part of the "dislocation and reassessment of our fundamental nature and role in the universe" (Floridi 2010, 12). As Floridi has pointed out:

> The criterion of existence—what it means for something to exist—is no longer being actually immutable (the Greeks thought that only that which does not change can be said to exist fully), or being potentially subject to perception (modern philosophy insisted on something being perceivable empirically through the five senses in order to qualify as existing), but being potentially subject to interaction, even if intangible. To be is to be interactable, even if the interaction is only indirect.
>
> (Floridi 2010, 12)

To be is to be interactable. This leads to an important consequence that is relevant from both a theoretical and a practical standpoint. We must revise our long-standing and traditional epistemological tendency to consider reality in terms of stable and enduring structures ultimately based on or reducible to the objective existence of a material, corporeal or physical reality. Our reality is increasingly the outcome of information-automated processes, software agents' behavior, delegated algorithmic decisions, intelligent ambient and smart technologies. However, virtual reality is no longer something we can only perceive and describe. It is primarily something we can interact with. Against this backdrop, the ontological question is no longer: what is virtual reality? The new ontological question is: what is virtual reality *for*? How can we interact with it? What functions can we attribute to it? How can virtual reality affect our experience of the world? The ontology of the virtual has itself become essentially technological, since what matters today is how we can interact with what we can build.

In this perspective, based on interaction, there are at least five constitutive elements of virtual reality: (1) *Agents*: most actions are no longer the exclusive prerogative of human beings. Artificial agents create their own reality and interact with human agents in virtual environments. (2) *Modes of reality*: different technologies can generate diverse forms of virtual reality (virtual reality; augmented reality; mixed reality; immersive technologies, etc.), with specific characteristics and problems. (3) *Entities*: virtual environments are populated by a rich variety of entities. The notion of entity must be understood as a broad category that includes objects; states of the world; events; and actions. (4) *Experience*: agents can undergo different experiences of virtual reality for different purposes: it is precisely these purposes that ultimately determine the type of experience that characterizes virtual reality. (5) *Consequences*: agents interacting with each other in virtual environments can produce real consequences. Increasingly crucial areas of study in fields such as law and economics are focusing on the real consequences of virtual action.

A full understanding of the relationship between technology and the ontology of the virtual would requires consideration of all five of these elements. Due to practical constraints, while some mention will be made of each, the main focus here will be on just one: experience.

3. Experience

Following a pragmatic approach, the focus here will be on the conceptual lens of experience. In fact, the core issue seems to lie in the type of experience that different agents engage in through the technological resources offered by virtual reality. The basic classification proposed here is far from capturing the full complexity of the problem. There are three ways of experiencing reality by means of the virtual: (1) augmentation/enhancement; (2) re-engineering; (3) evaluation/judgement. This classification was inspired by a conceptual scheme outlined by James Moor (1985), to which I have previously made recourse (Durante 2007, 2010). Moor's scheme is based on the types of questions that the development of technologies and their application to human tasks raise. This scheme can be summed up as follows.

The transformation of reality resulting from technological development is not just quantitative but also qualitative: the computer-based technological evolution not only widens the range of our interrogations, but also modifies the sense of our questioning. In other words, the meaning of a technological change is to be sought not only in the transformed reality (i.e., in virtual reality), but primarily in the transformation of the inquiry through which this reality is explored and cognitively represented. When reflecting upon a particular activity achieved by means of a computing device, we first ask ourselves (1) how efficiently the computer performs this activity. Then, when the technological application has become part of the performance of such an activity, we ask ourselves (2) what the nature and the value of this activity are. Finally, when the technology has, at least in part, changed the reality to which it applies, we wonder (3) whether this alternative state of the world can be used to reconsider or judge our previous condition. A brief example by Moor can illustrate the point: "For example, for years computers have been used to count votes. Now the election process is becoming highly computerized. Computers can be used to count votes and to make projections about the outcome. [. . .] The question is no longer 'How efficiently do computers count votes in a fair election?' but 'What is a fair election?' [. . .]. For better or worse, our electoral process is being transformed" (Moor 1985, 271). We are now ready to analyze our three categories of experience, now similarly transformed by virtual reality.

3.1 Augmentation/Enhancement

With the first category, we intend to refer to cases in which, through virtual reality, we can extend or enhance our previous experience of reality. In such cases, the effect on experience is quantitative rather than qualitative. New and further experiences are added to previous ones: the experience of flying a virtual airplane in a virtual world adds to the experience of flying a real airplane in the real world. However peculiar and unexpected, these experiences do not generally change our understanding of the entity, real or virtual, to which they apply. If we apply Moor's conceptual scheme, such experiences will lead us to formulate questions about how the entity is implemented, about the characteristics of a different mode of reality, about the way we interact with it, or about the real or virtual consequences of this entity, and so forth; however, they do not cause us to question the nature, meaning or value of the entity itself. This category is the most traditional one and includes the main forms in which virtuality can be technologically implemented: virtual reality (VR); augmented reality (AR); and mixed reality (MR).

Virtual reality (VR) is a popular term that has a well-documented history dating back to the mid-1980s. A recent book on the theme was written by Jaron Lanier (2017), the founder of VPL Research, the 1984 start-up that created the first VR commercial products. It contains more than forty definitions of VR, each of which describes a different aspect of this complex phenomenon. In a nutshell, there are at least three viewpoints from which to define virtual reality: (1) environment; (2) data; and (3) body.

From the viewpoint of environment, VR may be defined as a three-dimensional, computer-generated environment that can be explored and interacted with by a user. The more that users feel as though they are inside, or part of, a 3D computer-generated environment, the more they are immersed in that virtual world. In this sense, "VR replaces the real world altogether" and "places people inside virtual environments, letting them move around in it and interact with it as if it were the real world" (Lemley and Volokh 2018, 2). Moreover, VR can develop into, and be seen as, a Virtual World. The latter not only includes virtual reproduction of real entities but also enables them to evolve and propagate independently, mostly detached from our own reality, as is the case in digital environments such as Second Life or Entropia Universe, in which "fascinatingly, both endogenously produced economies and social orders" emerged (Bray and Konsynski 2007, 1).

From the viewpoint of data, VR may be defined as "an interactive network of information designed to connect data on a digitally created world. This digital world serves as an operating system to facilitate exchanges of information, and functions as a visual platform for applications and devices based on the infinite combination of data elements collected and shared by this virtual operating system" (Drovix 2009, 1). Needless to say, strong production, collection, sharing, and recombination of both personal and sensitive data make VR an economically disruptive technology, while raising legal issues of security, privacy and data protection, intellectual property, and safety (Fairfield 2012; Russo and Risch 2018; Thierer and Camp 2017; Lemley and Volokh 2018).

From the viewpoint of body, Jaron Lanier has stressed that "the visceral realness of human presence within an avatar is the most dramatic sensation I have felt in VR. Interactivity is not just a feature or a quality of VR, but the natural empirical process at the core of experience. It is how we know life. It is life" (Lanier 2017, 173). This gives us a further perspective on VR, as Lanier pointed out in an interview, because "The canvas of VR cannot be the external world—it has to be your body. An example of this is when you create out-of-body sensations of touch and feel. When you are really changing yourself, that is so much more interesting than watching something in the external world—and it really improves your sensation of reality" (Rubin and Lanier 2017). This is a powerful statement that reminds us that the impact of technology is both extrinsic, towards the world, and intrinsic, towards ourselves (Floridi 2014a, 87).

Augmented reality (AR) differs technologically and socially from virtual reality: "AR allows digital content to be layered over the real world. Using special glasses or, more commonly for now, a smartphone, AR users can see the real world as it actually exists, but with digital images superimposed on the world so that they seem to exist as part of the world. [. . .] Our experience of the real world will increasingly be overlaid with information and images" (Lemley and Volokh 2018, 2). Unlike virtual reality, AR is generally interoperable and is aimed more at enabling interactions with other users that may add to physical-world interactions, as with AR games such as Pokémon Go or Ingress. "In other words, AR supplements the natural environment users see around them; it does not completely replace the natural environment in the way that VR does" (Thierer and Camp 2017, 5).

Mixed reality (MR) integrates elements belonging to these two technologies (VR and AR) and mobile computing in the real world. MR draws on the fact that computing power is progressively entrusted to mobile computing that "takes computing out from behind the desk and into the real world" (Fairfield 2012, 64). "The central element of Mixed Reality is the tying of data to an anchor in the real world, be it a person, geographic location, or structure. [. . .] One new aspect that Mixed Reality introduces is the combination of mobile computers with geotagged data and the extent to which this combination is a part of everyday life. Through mobile devices, users see data that is tied up to particular places, objects, or people that they encounter. Smartphone technology and miniaturized computers permit a more mobile and interactive experience with our surroundings" (Fairfield 2012, 63–64). Via smartphones, smart glasses, or smart audio devices, MR applications can show diners review ratings as they walk by a restaurant, or enable soldiers on watch to receive realtime data about passing vehicles.

Let us sum up what has been said so far from the standpoint of experience: (1) VR is an experience of data-enriched virtual space; (2) AR is an experience of data-enriched virtual space that overlays real space; (3) MR is an experience of data-enriched real space. These technologies are not so much aimed at overcoming the dichotomy between the real and the virtual but rather aim to deepen it, to experience its possible combinations, and to exploit its wealth. The ontology of the virtual is thus expanded by its technological implementation, which allows us to have not only technologically mediated empirical (*visceral*) interactions but also *cognitive* experiences of different modes of reality.

Accordingly, let us examine what these technologies are mainly intended for. Their applications are wide-ranging and not just limited to gaming. This sort of technology can be applied in the fields of education (e.g., for documentary purposes); communication (e.g., for news reporting); art (e.g., interactive museums); movies and events (e.g., virtual theater); job training and system monitoring (e.g., for industrial simulation); healthcare (e.g., for treating post-traumatic stress disorder); engineering and product design (e.g., 3D renderings of objects); social life (e.g., for sex); driving (e.g., AR heads-up windshield displays); the military (e.g., for combat simulations), and so on. The list is longer than one might expect: "Digi-Capital predicts that AR and VR together will be a $150 billion business by 2020, with most of the revenue coming from outside of games" (Lemley and Volokh 2018, 10).

In all of these cases, technology expands or enhances some of our previous experiences of reality; they allow us to visit a museum in a different way or to see a stage play with 3D effects. They sometimes restore previous experiences, e.g., for people who have physical disabilities. In other instances, they stretch and challenge our habitual attitudes and behaviors, such as with an inspiring exercise of "diversity training" (Lemley and Volokh 2018, 8), where a business meeting can be held remotely using neutral avatars not defined by gender, age or race, thus minimizing prejudices (although this application may be also interpreted as actually erasing all signs of diversity rather than encouraging respect for diversity).

However, in some circumstances, technology may have a *re-engineering* effect, particularly when the consequences of actions in VR or AR confront us with uncharted problems, such as the hate speech phenomenon in virtual environments. As properly noted in this regard: "VR and AR will also challenge our understanding of what is speech (or, more precisely, communication)—and thus strongly protected by the First Amendment and other norms—and what is nonspeech conduct that merits regulation" (Lemley and Volokh 2018, 79). This leads us to the next category of virtual reality.

3.2 Re-engineering

With the second category, we intend to refer to cases in which virtual reality is used to re-engineer our former experience of reality (thus generating a new experience of reality). In such cases, the effect on experience is qualitative rather than quantitative. Re-engineering implies some fundamental transformation in the way we experience the world. To carry on with Moor's scheme, we are dealing with experiences that lead us to consider the nature, meaning and value of the activity to which they apply. The peculiarity of this category consists precisely in the fact that it causes us to rethink the nature, meaning and value of both real and virtual reality. The examples that follow should help illustrate this point.

The introduction of virtual currencies—whose earliest prototypes appeared in early Massively Multiplayer Online Games (MMOGs), such as Second Life and World of Warcraft—not only forces us to ask what a virtual currency is or to rethink the way in which certain financial transactions can be carried out; it leads us over time to wonder what currency will be like in the near future and what the financial economy will be like when currencies become virtual and financial transactions can be completely outsourced to artificial agents and algorithms (Coeckelbergh 2015). This does not imply, implausibly, that everything will totally change. Rather, it means that certain activities will at least require a profound change in perspective that needs to be better deciphered and regulated:

> While governments around the world observe the latest fintech developments, the regulatory approaches, like in the case of many emerging technologies, are notably lagging behind. The new digital economy, with the pace of its technological development, may require fundamental changes in our approach to regulation. Instead of ruling *ex post*, governments need to legislate *ex ante*, anticipating developments and preparing the regulatory landscape for robust readiness to meet continually evolving and accelerating challenges.
>
> (Caytas 2017, 4)

Consider another suggestive example, taken from the field of law: the case of indecent exposure in virtual spaces. This is a thought-provoking case since it questions what counts as harm. It also raises concerns about the extent of our own "freedom of sensescape" (Lemley and Volokh 2018, 63), when immersed in a digital environment (VR, AR, or MR). This requires us to consider why we should restrict things like indecent exposure when we do not restrain images of the same things elsewhere. Is virtual indecent exposure akin to the display of an image (speech) or to the threat of unwanted touching (conduct)? According to Mark Lemley and Eugen Volokh, it very well could be the latter, because "VR and AR, though, are deliberately created to make communicated image and sounds feel like real life. The technologies challenge our perception of the real because they blur the cognitive line between imaginary and physical presence" (Lemley and Volokh 2018, 80). In this sense, virtual indecent exposure might be considered in some circumstances as unprotected conduct constituting real harm. What is more important here is the lesson drawn by Lemley and Volokh as to the cognitive line between real and virtual reality:

> This in turn requires us to think seriously about some distinction we take for granted—between presence and remoteness, between speech and conduct, and between what is real and what is 'merely' perceived. If it turns out that the reason we ban indecent exposure is in part about perception and psychic harm rather than physical threat, that might cause us to rethink what it means to be hurt in a way that law cares about. If it turns out that we care about the perpetrator's intended behavior (and from his subjective perspective, his actual conduct) even in the absence of any harm to the victim, as we do in some but not all attempt law, that suggests a much broader notion of what we would punish if we only knew about it. And that has implications not just for the virtual world but also for the real world.
>
> (Lemley and Volokh 2018, 81)

It is no coincidence that the two examples given come from the fields of economics and law: they are both strongly institutionalized areas in which the real and the virtual are closely intertwined, and they are both constituted of social facts that are human constructs in a way that physical facts are not. This observation requires further comment. The distinction between these first two categories of experience draws on Philip Brey's distinction between simulation and ontological reproduction (Brey 2003) and Luciano Floridi's distinction between augmenting or enhancing technologies and re-engineering technologies (Floridi 2014a, 96–98). Let us briefly clarify this point.

First, we agree with Brey's account of virtual entities:

> Entities encountered in virtual worlds may be called virtual entities. At first glance, the ontological status of virtual entities is puzzling. [. . .] However, virtual entities are not just fictional objects because they often have rich perceptual features and, more importantly, they are *interactive*: they can be manipulated, they respond to our actions and may stand in casual relationship to other entities. So in our everyday ontology, virtual entities seem to have a special place: different from physical entities, but also different from fictitious or imaginary entities
>
> (Brey 2003, 276–277).

As already noted, interaction is the key conceptual lever with which to assess the ontological status of virtual entities. This allows virtual entities to differ from fictitious and imaginary entities. Virtual entities may thus produce and adjust their own reality. Therefore, a virtual world is not just the environment in which we encounter and experience virtual entities but is gradually becoming more of an environment where virtual entities can, at least in part, evolve and develop without our direct determination. Consider for example John McCormick and Adam Nash's work of art, *Reproduction*, an artificially evolving performative digital ecology, where there are "evolving virtual entities spawning and reproducing in virtual environments. "[. . .] The entities 'evolve,' 'reproduce,' 'live,' and 'die' over thousands of generations according to a constantly emergent evolution of those crude parameters that is informed, but not determined, by both their interactions with humans in the material world and with their interactions with each other" (Nash 2015, 22).

In his analysis of virtual entities, Brey also introduces a distinction between simulation and ontological reproduction that draws on John Searle's ontology of the real world (1995):[2]

> It seems, then, that there is a distinction between virtual entities that are accepted as mere simulations of real-world entities, and entities that are accepted as being, for all purposes, as real as nonvirtual entities. [. . .] So virtual versions of real-world entities are either mere *simulations*, that only have resemblance to real-world entities by their perceptual and interactive features, or *ontological reproductions*, which have a real-world significance that extends beyond the domain of the virtual environment. [. . .] Physical reality and ordinary social reality can usually only be *simulated* in virtual environments, whereas institutional reality can in large part be ontologically reproduced in virtual environments
>
> (Brey 2003, 277).

Although this is an interesting distinction that touches on several important issues, there are two reasons for which we shall depart from it in the rest of this chapter. Firstly, because we insist on *experience*, rather than *entity*, as the key conceptual lever of the ontology of virtual reality, and secondly, because we give a different meaning to simulation, as will emerge shortly, as a peculiar form of experience of virtual reality. Our distinction draws more on Floridi's categorization of augmenting and enhancing technologies as different from re-engineering technologies. Floridi's categorization is even more radical and far-reaching, since it involves the progressive blurring of the real/virtual dichotomy:

> The infosphere will not be a virtual environment supported by a genuinely 'material' world. Rather, it will be the world itself that will be increasingly understood informationally, as an expression of the infosphere. [. . .] We are changing our everyday perspective on the ultimate nature of reality from a historical and materialist one, in which physical objects and mechanical processes play a key role, to a hyperhistorical and informational one. This shift means that objects and processes are *de-physicalized*, in the sense that they tend to be seen as support-independent.
>
> (Floridi 2014a, 50)

In addition to this progressive blurring of the line between the real and the virtual, we will be confronting a less apparent but even more relevant consequence: the need for different *epistemic* agents to adapt their representation and experience of the world to their own peculiar way of acting and functioning. For example, the main issues regarding the difference between self-driving cars and cars driven by human beings are generally considered today in terms of moral, legal or technological questions. Obviously, these issues exist and are here to stay. However, the true confrontation is going to be epistemic. It will concern the way in which different agents determine which road to take, which obstacles to recognize and which to overtake, in short, it will concern agents' knowledge of the reality in which they operate. Who will adapt what representation of the world to whom? Even the legal consequences in the real world will mostly depend on the epistemology with which the agents represent the world and act within it.

I must reiterate that for artificial agents, software systems, or algorithms, the reality that is represented in the virtual format is the only reality that exists. One of Floridi's central claims is that we are likely to gradually adapt our environment to the representations and configurations of the world that are instrumental to the functioning of the computational systems we have created and let loose:

> You need to adapt the environment to the robot to make sure the latter can interact with it successfully. Likewise, put artificial agents in their digital soup, the Internet, and you will find them happily buzzing. The real difficulty is to cope, like the wasps, with the unpredictable world out there, which may also be full of traps and other collaborative or competing agents.
>
> (Floridi 2014a, 136)

These first two categories point in a single direction: technologies of the virtual tend to alter our experience of reality. They do so insofar as they expand upon or improve it, as they change its meaning or nature, and finally, as we will see shortly, in what they evaluate or how they judge it, by imagining an alternative version of reality. Let us now turn to this third category: evaluation/judgement.

3.3 Evaluation/Judgement

With the third category, I refer to those cases in which, through virtual reality, we can evaluate and judge our former experience of reality from a different perspective. This is made possible through the use of simulations. In the context of virtual reality, the term simulation usually refers to the reproduction of real objects in a virtual format. Here we are using it in a different sense[3], to refer specifically to the algorithmic production of counterfactuals.

Consider the 2016 movie "Sully" by Clint Eastwood. The plot is based on a true story, although it is of no concern to us here how faithful the story was to reality. The events recounted in the film are as follows. In January 2009, US Airways flight 1549 smashed into a flock of geese, three minutes from New York City's LaGuardia airport. With both engines out, pilot Chesley "Sully" Sullenberger, assisted by co-pilot Jeffrey Skiles, made the decision not to attempt to reach an airport runway. Instead, Sully dead-sticked his Airbus 320 to a landing in the Hudson River, saving the lives of 155 people. Later, the National Transportation Safety Board claimed that several confidential computerized simulations indicated that the plane could have landed safely at the LaGuardia or Teterboro Airports without engines. In the end, Sully managed to prove that the computerized simulations were wrong, and that there were no real alternatives to what he had done.

This, however, is not the point. The point is that simulation can be used not only to reproduce a real object in a virtual format, as with augmented virtual reality, for training pilots. The simulation is rather used to scrutinize events that have already occurred and to evaluate what could have happened otherwise. A computer simulation is used to produce what we might define as an *algorithmic counterfactual*, or in other words, a representation of reality in a virtual world that allows us to assess and judge what has happened in the real world. The algorithmic counterfactual actually works to tell us how the agent should have behaved in reality.[4]

I have decided to isolate this meaning of simulation and to make it a category in its own right, for it represents, in my view, an area of great potential for the expansion and future development of virtual reality. It is here—in the algorithmic production of counterfactuals—that technology and the ontology of the virtual are likely to be fused. Once again, this is not a completely novel idea: we have always nourished the imagination of alternative and idealistic worlds as a viable criterion of counterfactual judgment. At the same time, however, the notion has undergone substantial changes. For instance, the virtual reality of an algorithmic counterfactual is profoundly different from a Kantian regulative ideal of reason. While the former tells us how things could or should have been in the proximate alternative world, the latter only expresses archetypal standards, which can be approximated but never fully attained. Those regulative ideals or archetypal standards assign a direction to a practice, but do not rule out practices.

When one has an idea—a promising idea—it is important to ask whether someone else has already thought of it. That, of course, turned out to be the case here. While I was pleased to have come up with the idea of an algorithmic counterfactual, I then discovered that Judea Pearl and Dana McKenzie (2018) have also recently discussed the algorithmization of counterfactuals.

While I am worried about the whole chain of problems concerning the use of algorithms[5] (from potential cognitive biases that affect the formation of the knowledge base to the lack of transparency of the black box[6] of an inference engine, up to the possible discriminatory effects of the algorithmic outcome), Pearl and McKenzie attribute a decisive and beneficial role[7] to the outsourcing of algorithmic counterfactuals to thinking machines:

> Counterfactuals are the building blocks of moral behavior as well as scientific thought. The ability to reflect on one's past actions and envision alternative scenarios is the basis of free will and social responsibility. The algorithmization of counterfactual invites thinking machines to benefit from this ability and participate in this (until now) uniquely human way of thinking about the world.
>
> (Pearl and McKenzie 2018, 10)

Pearl and McKenzie are completely aware of the delegation problem ("This brings up a natural question: How much can we trust the computer simulations?" [2018, 295]). What puzzles me about the normative function of algorithmic counterfactuals is that their ability to predict alternative worlds is mostly based, ultimately, on a static training data set describing the world as it is. The algorithmic what-ifs of the most proximate worlds tend to reaffirm the primacy of the existent in a more or less surreptitious manner. In this way, the virtual is not instrumental to the imagination of different worlds, but rather to the prediction of expected and desired worlds. This raises a serious issue that warrants discussion in the near future: the ontology of the virtual will primarily be technologically determined by "prediction machines" (Agrawal, Gans, Goldfarb 2018), given that prediction is at the heart of our information societies. Although often unseen, a critical area of development in virtual reality is one in which machines and human beings predict behaviors, decisions and other expected results through simulations, through which reality can then be judged and modified. It will increasingly be the case that these predictions will serve as the basis, for example, for recognizing or discarding rights or other prerogatives.

Let us now turn our attention from experience to some aspects of the real consequences of virtual reality.

4. Consequences

A prominent feature of virtual reality systems is their potential to have consequences in real life. As already noted, fields such as law and economics are privileged areas where it is possible to appraise the real consequences of virtual reality. In virtual environments, we can carry out a series of actions that can have ramifications in real life, such as making legally relevant agreements, committing crimes, following rules of conduct, establishing governance, making economic transactions, trading convertible currencies, selling items, paying taxes, and so forth. The interested reader is invited to turn to the extensive literature on the topic. Here we will focus our attention briefly on three general aspects that affect the real consequences of virtual activities: (1) typification; (2) ontological marks; and (3) data.

4.1 Typification

The combination of digitization and virtualization has generated not only an increasing process of dephysicalization but also of the typification of people, processes, and objects, as Luciano Floridi has observed:

> When our ancestors bought a horse, they bought *this* horse or *that* horse, not 'the' horse. Today, we find it utterly obvious and non-problematic that two cars may be virtually identical and that we are invited to test-drive and buy the model rather the individual 'incarnation' of it. We buy the type not the token. [. . .] Quite coherently, we are quickly moving towards a commodification of objects that considers repair as synonymous with replacement.
>
> (Floridi 2014a, 57)

This means that not only objects or processes but also human beings no longer count as unique and irreplaceable entities but as instances of a type, in which a general rule of substitutability prevails. This process of typification has therefore generated two related consequences, which are particularly relevant with regard to virtual reality:

> Such a shift in favor of types of objects has led, by way of compensation, to a prioritization of informational *branding*—a process comparable to the creation of cultural accessories and personal philosophies—and of *reappropriation*. [. . .] the processes of dephysicalization and typification of individuals as unique and irreplaceable entities may start eroding our sense of personal identity as well. We may risk behaving like, and conceptualizing ourselves as, mass-produced, anonymous entities among other anonymous entities, exposed to billions of other similar individuals online. We may conceive each other as bundles of types, from gender to religion, from family role to working position, from education to social classes.
>
> (Floridi 2014a, 57–58)

People have begun to conceive of themselves as virtual bundles of selectable, modifiable or improvable properties and characteristics in order to create a unique self-representation that is appropriate to one's own expectations or, more often, to social ones: "the dialectics of being uniquely like everybody else joins forces with the malleability of the digital to give rise to the common phenomenon of 'airbrushing.' Digital photographs are regularly and routinely retouched in order to adapt the appearance of portrayed people to unrealistic and misleading stereotypes, with an unhealthy impact on customers' expectations, especially teenagers" (Floridi 2014a, 57). The sense of loss or erosion of personal identity is heightened by the fact that, in digital and virtual environments, this is increasingly the result of the chief and growing importance in machine learning of training the models to learn and apply class labels properly, to sort all things in a real or virtual environment into the proper class bucket.[8] Actual references (i.e., direct experiences) and ontological marks (i.e., what signals the difference between real and virtual worlds) are gradually vanishing, raising the second issue which generally affects the consequences of virtual reality.

4.2 Ontological Marks

Immersive technologies—which include versions of VR, AR, and MR—are specially designed to mimic reality as closely as possible, so as to blur the line between physical and virtual worlds. Such technologies tend to remove the ontological marks that make users perceive the distinction between the real and the virtual in order to increase the feeling of immersion. Immersive technologies are characterized by problems that generally affect data-intensive virtual technologies. These problems may concern legal issues such as privacy, intellectual property, security, and physical integrity; psychological issues such as isolation, distraction, anxiety, and addiction; and social and moral issues such as aggression, hate, contempt, and racism. Immersive technologies however may also raise the following issue:

> But if you see an avatar in a VR world, you are seeing it in a context specially designed to mimic reality as much as possible. When you turn your head, the illusion created by VR is reinforced, not broken. In more advanced VR systems, you might be walking around on a two-dimensional treadmill rather than just sitting in your armchair. Moreover, you will see the avatar not in some special context that you bring up just to see impersonations [. . .]. Rather, you might see the avatar in your ordinary 'travels' in the VR environment. Even if you logically recognize that the avatar is a pseudonym, it will feel like a person.
>
> (Lemley and Volokh 2018, 69)

The blurring of the ontological marks of reality, in an immersive technology, can therefore weaken or exclude defenses or justifications for reprehensible actions or attitudes, based on the alleged perception of the difference between the real and the virtual. In other words, the stronger the perception of reality, the more difficult it will be to invoke the virtual dimension of the context as a justification for (the real consequences of) our actions (consider, for instance, the case in which someone insults or is aggressive towards an avatar that is felt as a real person). What kind of response does this issue evoke, together with the concerns highlighted earlier? What kind of governance of immersive technologies should we adopt?

Adam Thierer and Jonathan Camp have provided some possible answers, highlighting two alternative governance visions that could govern the future of immersive technology:

> *Precautionary principle reasoning* refers to the belief that new innovations should be curtailed or disallowed until developers can demonstrate that the innovations will not cause any harm to individuals, groups, specific entities, cultural norms, or various existing laws or traditions. The alternative vision of *permissionless innovation* refers to the idea that 'experimentation with new technologies and business models should generally be permitted by default. Unless a compelling case can be made that a new invention will bring serious harm to society, innovation should be allowed to continue unabated and problems, if they develop at all, can be addressed later.'
>
> (Thierer and Camp 2017, 27)

Without going into detailed analysis of these alternatives, I agree that: "to make permissionless innovation the basis of public policy toward immersive technology, policymakers should adopt the following 10-part blueprint" (Thierer and Camp 2017, 33), according to which a detailed policy framework[9] is elaborated on the basis of the paradigm "educate and empower," instead of that of "legislate and regulate" (Thierer and Camp 2017, 42). Needless to say, innovative immersive technologies should be examined and discussed at least through the conceptual lens of "pro-ethical design," which operates at the informational and not at the structural level of a choice architecture (Floridi 2016). For the sake of brevity, we can say that the ultimate and most essential reason for favoring a more permissive attitude is to be found in what has been properly observed by Ithiel de Sola Pool, with reference to the regulation of information markets:

> Enforcement must be after the fact, not by prior restraint [. . .]. Regulation is a last recourse. In a free society, the burden of proof is for the least possible regulation of communication.
>
> (de Sola Pool 1983, 231)

This being said, some implications of the disappearance of ontological marks may entail normative issues that require some form of regulation. Consider, for instance, the case of Google's Duplex launch of an AI-based voice that takes on human vocal tics (uptalk, interjection, hesitation, etc.) to produce a seamless, ontologically indistinct experience of a virtual assistant calling on behalf of a real agent. Such mimicry can foster trust in the service while simultaneously raising concerns about deception. More broadly, the phenomenon of synthetic media (synthetic text-to-speech technology, deepfakes, GAN-generated faces, etc.) may raise concerns about the epistemic trust of people in the daily experience of (less and less perceptibly fake) messages and images, which can generate morally, politically, and legally relevant consequences. This may require synthetic media "watermarks" that will allow for the distinction between real and fake to be preserved. Finally, let us turn to the third general aspect concerning the real consequences of virtual reality, which regards the fact that virtual reality technologies are data-intensive technologies.

4.3 Data

Virtual reality technology, like most information and communication technologies, is based on data: it collects, stores, produces, and shares data. This includes all kinds of data, including the personal and sensitive. Users immersed in a virtual environment are part of a context that is built entirely of data. Data do not necessarily belong to—or remain at the disposal of—the users who produce them:

> Our movements and actions in the physical world are increasingly observed, recorded, and tracked. But there are still spaces where we are not followed and acts that are not recorded and searchable. In VR that will likely not be true. Everything we do, we do before an audience—a private company that may well keep and catalog that data, and may have lots of reasons to do so (data mining, security, user convenience, and more).
>
> (Lemley and Volokh 2018, 17)

This raises a critical issue with several consequences in real life: that is, "the ownership and control of data" (Agrawal, Gans, and Goldfarb 2018, 174). Private companies—which provide hardware and software resources for VR—record, collect, and store users' data, with the result that "those private companies will invariably impose terms of use that purport to bind users of the hardware and software. Those terms may disclaim liability for harm. They may assert ownership over the things we create in VR. And they may require us to consent to having information about our conduct in the virtual world recorded and shared" (Lemley and Volokh 2018, 17).

Data is needed not only to enhance and develop virtual reality technologies but also and above all to fuel AI systems, machine learning, prediction machines, and to profile users for commercial purposes. Virtual reality may thus become—and partly already has become—a gigantic repository to draw on, in order to mine data-patterns, predict behaviors, and understand trends, in order to suggest choices, influence actions, and allocate resources. This reinforces a trend that began with the Internet and has continued with many mobile applications: users have limited contractual powers. While the consensual paradigm based on notice and consent ensures the application of contract law, it hardly protects the rights and freedoms of data subjects. Users of virtual reality technologies cannot follow the entire life-cycle of their own data.

Ownership and control over collected data is not the only issue related to data-intensive technology of virtual reality. There is also the significant issue of data reliability. Users of virtual reality technologies have nothing but the data to back up their decisions and actions. The data must therefore be as accurate as possible in order to allow users to prudently make decisions and adopt behaviors with potential consequences in real life. Needless to say, problems of epistemic trust—regarding which data we can rely on when deciding and acting—also concern the real world. However, in reality, users can rely on direct experience, validation, tangible physical properties, ontological marks, content asserted by epistemic authorities, in short, on filters of reliability, which in virtual reality are more evanescent. While there are many tests for verifying and validating the reliability of VR systems from a technological standpoint, it is still unclear how data accuracy and reliability can be granted to users when immersed in virtual environments. However, this problem is open and will require discussion, since the accuracy and reliability of data are an integral part of the necessary security and safety of virtual environments.

5. Conclusions

Digital technology generates the new ontology of the virtual. It is pointless and perhaps even counterproductive to try and draw a line between physical and virtual reality in strict ontological terms, for it is we—through technological innovation—who move and reshape that line. It is thus more relevant and fruitful to ask in pragmatic terms how the technology of the virtual modifies the way we experience reality. In doing so, we find that it happens in at least three ways.

First of all, the technology of the virtual allows us to expand or improve our experience of reality. This occurs in many ways: by recovering or compensating for a lost capacity; by providing new mental or bodily experiences; or by gaining a novel perception of reality from a totally different point of view. It can also occur to the benefit of our own experience of physical reality. In this respect, Jaron Lanier is right: "A coarser, simulated reality fosters appreciation of the depth of physical reality in comparison. As VR progresses in the future, human perception will be nurtured by it and will learn to find even more depth in physical reality" (Lanier 2017, 50).

Second, the technology of the virtual can affect our experience of reality even more drastically, in modifying and re-engineering the world to which it applies. This alters our understanding and representation of both reality and ourselves. What is important to grasp in this regard is that we human beings are—as agents—part of that world that technology modifies and re-engineers. This means that we do not merely interact only with different modes of reality (VR, AR, or MR), but also with other epistemic, artificial agents that produce their own representations of reality, which mingle with ours: mixed reality is not so much a technological platform as an intermingling of the cognitive and the epistemic.

Finally, the technology of the virtual can also serve another purpose: that of constructing a counterfactual reality that judges our reality and drives us to amend it. Simulation is thus not limited to enabling us to have new or different experiences of reality, but allows us to compare alternative experiences and to reflect on how reality should have been. This is the case in which the ontology of the virtual—through the algorithmic construction of counterfactuals—acquires a normative function. Despite the heuristic value of counterfactual simulation that we recognize, we believe that this normative dimension requires attention, public discussion, and criticism, as in every circumstance in which norms are produced or conveyed outside a deliberative process (Zittrain 2007; Pagallo and Durante 2016).

In conclusion, interaction stands out as the ontological criterion of existence and as the main factor of the technological advancement of virtual reality. Hence, we consider the virtual as part of the real and the real as part of the virtual. This tells us something crucial. We no longer merely describe reality and act, accordingly, on the basic of such a description. We are constantly constructing the reality—whether virtual or real—in which and with which we interact. Since this ability grows out of the interaction process, it is impossible to determine in advance or as an immutable certainty what is to be taken as real and what as virtual. We move, blur and reshape this dividing line. We also cross it and, in so doing, extend our understanding of it as virtuality expands, enhances, re-engineers or subjects to criticism and amendment our own experience of reality.

Notes

1. For more on this perspective, see Floridi (2014, 238).
2. On the debate about cyberspace, ontology, and virtual reality between Platonic dualism and Searlean realism see also, respectively, Heim (1993; 1998) and Koepsell (2000).
3. The potential of virtual reality simulation has also been studied, in order to address implicit racial basis among judges and jurors in the courtroom setting. On this see Salmanowitz (2016 and 2018).
4. In a sense, simulation can show even more than how the agent "should" have behaved: it can show how the agent *did* behave (reasonably, skillfully, responsibly, or not) and how the real world actually was (what possibilities actually existed, what elements of witness testimony to the reality could or could not have occurred in fact). Therefore, algorithmic simulation may be also evidence of the *factual*.
5. See D'Agostino and Durante (2018).
6. See Pasquale (2015).
7. Furthermore, the role of counterfactuals as a criterion for explaining an automated decision has recently been highlighted by Wachter, Mittelstadt, and Russell (2018).

8. Ted Striphas (2015) speaks in an evocative way of “algorithmic culture” to signal that the massive recourse to algorithms does not only affect numerous decision-making processes but also our culture, generating forms of reliance on new ways of representing the reality.
9. For more on this point, see Thierer (2016).

References

Agrawal Ajay, Joshua Gans, and Avi Goldfarb. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard: Harvard Business Review Press.

Bray, David, and Benn Konsynski. 2007. “Virtual Worlds: Opportunities for Multi-Disciplinary Research.” *The Data Base for Advances in Information Systems*, Special Issue on Virtual Worlds, 38(4): 1–18.

Brey, Philip. 2003. “The Social Ontology of Virtual Environments.” *American Journal of Economics and Sociology*, Special Invited Issue: John Searle’s Ideas about Social Reality: Extensions, Criticisms, and Reconstructions, 62(1): 269–282.

Caytas, Joanna. 2017. “Regulatory Issues and Challenges Presented by Virtual Currencies” (May 30, 2017). *Columbia Business Law Review*. Available at SSRN: https://ssrn.com/abstract=2988367.

Clauberg, Johannes. 1647. *Elementa philosophiae seu Ontosophia*. Groningen.

Coeckelbergh, Mark. 2015. *Moral Machines. Electronic Financial Technologies, Distancing, and Responsibility in Global Finance*. London: Routledge.

Courtine, Jean-François. 1990. *Suarez et le système de la métaphysique*. Paris: PUF.

D’agostino, Marcello, and Massimo Durante, eds. 2018. “The Governance of Algorithms.”. *Philosophy & Technology*, Special issue on the Governance of Algorithms, 31(4): 499–653.

De Sola Pool, Ithiel. 1983. *Tecnologies of Freedom*. Boston: Harvard University Press.

Drovix Pascal. 2009. “The Reality of Virtual Reality” (April 19, 2008). Online published (September 15, 2009). Available at SSRN: https://ssrn.com/abstract=1473031.

Durante, Massimo. 2007. *Il futuro del web: etica, diritto, decentramento. Dalla sussidiariet*à *digitale all’economia dell’informazione in rete*. Torino: Giappichelli.

Durante, Massimo. 2010. “Re-designing the Role of Law in the Information Society: Mediating between the Real and the Virtual.” *European Journal of Legal Studies* 2(3): 1–18.

Fairfield, Joshua. 2012. “Mixed Reality: How the Laws of Virtual Worlds Govern Everyday Life.” *Berkeley Technology Law Journal* 27: 55–116.

Floridi, Luciano. 2008. “The Method of Levels of Abstraction.” *Minds and Machines* 18(3): 303–329.

Floridi, Luciano. 2010. *Information. A Very Short Introduction*. Oxford: Oxford University Press.

Floridi, Luciano, ed. 2014. *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Dordrecht: Springer.

Floridi, Luciano. 2014a. *The Fourth Revolution. On the Impact of Information and Communication Technologies on Our Lives*. Oxford: Oxford University Press.

Floridi, Luciano. 2014b. “The Latent Nature of Global Information Warfare.” *Philosophy & Technology* 27: 317–319.

Floridi, Luciano. 2016. “Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration.” *Science and Engineering Ethics* 22(6): 1669–1688.

Gilson, Etienne. 1952. *Being and Some Philosophers*. Toronto: Pontifical Institute for Mediaeval Studies.

Göckel, Rudolph. 1613. *Lexicon philosophicum quo tanquam clave philosophiae fores aperiuntur*, Marburg (reprint: Hildesheim: Georg Olms, 1980).

Harari, Yuval. 2015. *Sapiens: A Brief History of Humankind*. New York: HarperCollins.

Heim, Michael. 1993. *The Metaphysics of Virtual Reality*. Oxford: Oxford University Press.

Heim, Michael. 1998. *Virtual Realism*. Oxford: Oxford University Press.

Koepsell, David. 2000. *The Ontology of Cyberspace*, Chicago: Open Court.

Lanier Jaron. 2017. *Dawn of the New Everything: A Journey Through Virtual Reality*. New York: Bodley Head.

Lemley, Mark, and Eugene Volokh. 2018. "Law, Virtual Reality, and Augmented Reality" (February 27, 2018). *University of Pennsylvania Law Review*, 166: 1–82.

Lorhard, Jacob. 1606. *Ogdoas Scholastica, continens Diagraphen Typicam artium: Grammatices (Latinae, Graecae), Logices, Rhetorices, Astronomices, Ethices, Physices, Metaphysices, seu Ontologiae*. Sangalli: Apud Georgium Straub.

Moor, James. 1985. "What Is Computer Ethics?" In *Computers & Ethics*, edited by Terrel Ward Bynum, 266–275. Malden: Blackwell Publisher.

Nash, Adams. 2015. "An Aesthetics of Digital Virtual Environments." In *New Opportunities for Artistic Practice in Virtual Worlds*, edited by Denis Doyle, 1–22. Hershey: IGI Global Publisher.

Øhrstrøm Peter, Henrik Schärfe, and Sara L. Uckelman. 2008. "A 17th Century Hypertext on the Reality and Temporality of the World of Intelligibles." In *Conceptual Structures: Knowledge Visualization and Reasoning*, edited by Peter Eklund and Ollivier Haemmerlé, 74–87. Proceedings of the 16th International Conference on Conceptual Structures, ICCS 2008 Toulouse, France, July 7–11. Dordrecht: Springer.

Pagallo, Ugo, and Massimo Durante. 2016. "The Pros and Cons of Legal Automation and Its Governance." *European Journal of Risk Regulation*, 7(2): 323–334.

Pasquale, Franck. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Boston: Harvard University Press.

Pearl, Judea, and Dana McKenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

Rubin, Peter, and Jaron Lanier. 2017. "A Conversation with Jaron Lanier, VR Juggernaut." *Wired* 21.11, available at https://www.wired.com/story/jaron-lanier-vr-interview/.

Russo, Jack, and Michael Risch. 2018. "Virtual Copyright." In *Research Handbook on the Law of Virtual and Augmented Reality*, edited by W. Barfield and M. Blitz. Cheltenham: Edward Elgar Publishing. Available at SSRN: https://ssrn.com/abstract=3051871.

Salmanowitz, Natalie. 2016. "Unconventional Methods for a Traditional Setting: The Use of Virtual Reality to Reduce Implicit Racial Bias in the Courtroom." *University of New Hampshire Law Review* 15(1): 117–160.

Salmanowitz, Natalie. 2018. "The Impact of Virtual Reality on Implicit Racial Bias and Mock Legal Decisions." *Journal of Law and the Biosciences* 5(1): 174–203.

Searle, John. 1995. *The Construction of Social Reality*. Cambridge, Boston: MIT Press.

Striphas, Ted. 2015. "Algorithmic Culture," *European Journal of Cultural Studies*, 18(4–5): 395–412.

Thierer, Adam. 2016. *Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom*. 2nd ed. Arlington: Mercatus Center at George Mason University.

Thierer, Adam, and Jonathan Camp. 2017. "Permissionless Innovation and Immersive Technology. Public Policy for Virtual and Augmented Reality." *Mercatus Working Paper*, Mercatus Center at George Mason University, Arlington, 1–52.

Wachter Sandra, Brent Mittelstadt, and Chris Russell. 2018. “Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR.” *Harvard Journal of Law & Technology*, 31(2): 1–51.

Zittrain, Jonathan. 2007. “Perfect Enforcement on Tomorrow’s Internet.” In *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes*, edited by Roger Brownsword and Karen Yeung, 125–156. London: Hart.

CHAPTER 17

USING PHILOSOPHY OF LANGUAGE IN PHILOSOPHY OF TECHNOLOGY

MARK COECKELBERGH

1. Introduction: Thinking about Language and Technology as a Response to the Empirical Turn

After the empirical turn in philosophy of technology (Achterhuis 2001), which tried to get away from twentieth century abstract metaphysical philosophies of technology by focusing on our embodied and material engagements with technology, authors such as Don Ihde and Andrew Feenberg have focused on understanding and evaluating technological artifacts. This is and has been a very fruitful route of inquiry, but the focus on material artifacts has been at the expense of neglecting the roles language plays with regard to technology (Coeckelbergh 2017a, 2017b). One reason why empirical philosophers of technology such as Ihde and people such as Bruno Latour turned away from thinking about language is that some twentieth century philosophies of language were too concerned with the abstract symbolical, for example in postmodernism. The rejection of postmodernism's obsession with signs is understandable. But the neglect of theory about language and in effect neglect of insights from an entire subfield of philosophy—philosophy of language—is problematic if philosophy of technology is to do justice to the linguistic dimension of technology and technology use, while also developing into a mature discipline that interacts with other subfields in philosophy.

This chapter aims to contribute to further articulating and filling this gap by showing *some* ways in which language and technology are connected, and by making some links between philosophy of language and philosophy of technology. This chapter is not intended to offer a comprehensive integration of these subfields but to explore some potential bridges; in particular, bridges to Wittgenstein, Ricoeur, and Searle. In order to clarify my arguments and interpretations, I will use examples from robotics and other information and communication technologies throughout the chapter.

The chapter consists of three sections. The first section shows (a) how language is often literally (note the text-based metaphor) connected to technology, for example in contemporary information and communication technologies such as software and assistive devices; to adequately describe their ontology and agency, referring to language seems essential. Talking about material artifacts will not suffice for philosophers of technology; the linguistic dimension needs to be taken into account. Moreover, (b) the use of words and the use of things often go hand in hand. For example, if I use applications such as Skype or Whatsapp to talk with someone via my phone, I am using words and a technological artifact (the phone) in one and the same act. Technology and language also align in the use of intelligent assistive devices such as Alexa, a voice interface embedded in an artifact. In such examples, however, language and technology are understood as mere instruments.

The second section outlines some ways in which language plays a role that is not merely instrumental. It argues that (a) the discourse about technologies influences the development and use of technologies and that (b) language plays a mediating role in technological practices and concrete human-technology relations and interactions. The latter claim is a response to postphenomenology's insistence that material artifacts are mediators, which usually does not consider *other* mediators, and a response to Searle's social ontology. A suggestion is made for how to integrate the mediating role of language into Ihde's postphenomenology of human-technology relations.

The third section argues that language and theory about language can also be used as a model and metaphor for understanding technology. It gives three examples of how one could use notions from philosophy of language (broadly conceived) in philosophy of technology: language games and form of life (Witttenstein) and narrative (Ricoeur). Here the point is not to say something about language, but rather to use similar theoretical notions to understand and evaluate technologies. These concepts borrowed from philosophy of language offer insights into the holistic and temporal-narrative dimensions of technology use.

The result is a palette of options for the further study of the relation between language and technology, and indeed for the further use of philosophy of language in philosophy of technology. Of course this overview is not meant to be exhaustive; there may be many more ways in which language and technology are connected, and there are of course many more theories and approaches in philosophy of language that could be used. This is just a selection to show how it could work. Moreover, by making connections between the material and the linguistic dimension in technologies, the chapter also constitutes a critical response to the empirical turn in philosophy of technology, in particular to postphenomenology: this chapter is to be read as an attempt to redress the latter's overemphasis on the material aspect of technology to the neglect of understanding language as a mediator.

2. Language, Literally: How Language Is (very often) Ontologically and Pragmatically Entangled with Technology

Language is often literally connected to technology, or is even part of what the technology is. Consider some of the technologies most of us use daily, such as computers and mobile phones (or less common technologies such as robots). These are material artifacts, to be sure, but they run on software, which is based on a software language, on code. Both the material aspects and the linguistic aspects constitute the technology. For the technology to work, it is crucial that they are connected. Code by itself cannot do anything in the world; the hardware (and other software) is needed. Vice versa, the hardware needs the language of software to do things. When it comes to contemporary information and communication technologies (ICTs), it is clear that language and materiality are entangled. The agency and ontology of such technologies cannot be adequately described without taking language into account. Even at the so-called technical level, these technological artifacts already have a hybrid nature: they are material and linguistic at the same time.

Moreover, when we consider *use* and interaction with ICTs, in many cases language is crucial since it is a key part of the interface between the human user and the technology. This was already true for personal computers and the Internet, which use text-based interfaces, and it is even more true for so-called social media and for new social devices such as home assistants and social robots, which increasingly use voice-based interfaces. Both text and voice communication rely on language. There are material devices such as a computer with keyboard or a robot, but the actual use of, and interaction with, the use of device is mediated by language. Moreover, language also mediates interaction with others, through the technology. Phenomenologically speaking, the user is not interacting with the material device. The user is either communicating and interacting with other users or with the "personality" of the device (e.g., an assistive device or robot), and both language and material technology mediate and make possible this communication and interaction.

Thus, language is part of technology (e.g., a computer always includes code to function) or simply *is* a technology; for example, an interface technology, a communication technology, an information technology. It is an instrument to interface between human and material hardware, it is an instrument to talk with people, it is an instrument to share information, and so on. This is also true for older ICTs. Think about a book, for example: ontologically, it is both matter and language; there is a material dimension and a sign dimension. There is no such thing as a text without a book or another material carrier (e.g., the screen of your mobile phone), and a book without text is not a book (it might be a work of art, for example, which draws attention to the materiality of the medium).

Furthermore, the book is a means of communication, for example, communicating information about a subject or communicating a narrative. It is also an interface between author and reader and between readers. It is a technology, and it has both material and linguistic dimensions.

Now one may object that all this is true for ICTs but not for tools such as a hammer. A hammer seems not ontologically connected to language (language or text is not part of what the hammer is) and its use does not depend on language; I do not have to use words to use the hammer. This is true, and one could already conclude that attention to language can enable us to distinguish between different kinds of technologies: some are more language-based and language-dependent than others. Note also that this is one way of showing that attention to language does not mean that one loses focus on the concreteness of technologies, as philosophers of the empirical turn may fear. Instead, attention to language invites us to describe more precisely the particular concrete ways in which technologies and language are entangled.

But this is not the end of the story: it must be acknowledged that (a) we often *do* use words when we use things such as hammers, since often use of technology is collaborative and requires communication. When we think about what we do we also often use language (perhaps always—but this is controversial among philosophers). And, related to the latter point, (b) the meaning of "hammer" for us, humans, is always a meaning that is linguistically (and socially) mediated. The thing has a name, and when we use a hammer, we also use a word (at the same time). Name and thing are connected in use. We cannot normally even *think* of a hammer without using the word hammer, or an equivalent linguistic sign for one. As a child, we learn what a hammer is, and this means that we make the connection between name and thing (and between thing and use), a connection which afterwards remains tight. It also means that both word and thing are part of a social-linguistic community in which this word and this technology make sense (i.e., are used in this way). One could object that the word "hammer" is a mere representation of the actual material hammer. But this supposes that we can conceive of the material hammer without using the word or *a* word. As skilled language users, however, we normally can no longer see or use the hammer without seeing or using the word. We can use a different word, perhaps (e.g., from a different language or we can make up a word and share that meaning with others), but there will be always a *word* connected to a thing.[1]

Another way of saying this is that words do not just represent material things, they are not mere tools for classifying physical objects. They are perhaps not necessarily connected to particular things (that is, in theory there is no necessary relation), but as Heidegger and Wittgenstein already suggested, *in practice* we already find ourselves in a world full of word-things, and indeed a world full of meaning. One could say: words and things are connected in use, not in theory (I will say more about Wittgenstein later in this chapter). But if words are not (mere) representations or tools, then this raises the question of whether language can be more than just an instrument (e.g., for communication or for representation), more than a neutral mediator, and more than something that is external to technology. Let me try to conceptualize this in the next section.

3. Language as More than an Instrument: How Words Shape What Technology Is

Words matter. Names matter. They are not neutral but shape what the thing is. We do not merely ascribe words to things. Searle argued in his social ontology (1995; 2006) that what renders things social is that we use words that perform a declaration and in this way give meaning to things. For example, paper money is only money because we declare (and agree) that it is money. Searle thus made a sharp distinction between, on the one hand, the material artifact, and on the other hand, the linguistic act (speech act). Against Searle, one could argue that the thing is already meaningful at the moment we use it and what the thing is cannot be disconnected from language and its use within a particular social-linguistic community. In contrast to Searle, one could argue that the social meaning is not ascribed to the meaningless thing, but rather that the thing is already socially meaningful through language (and the language community).

Another way of saying this is to claim that artifacts are linguistically constructed. As I have argued in an article about language and robots (Coeckelbergh 2011), how we linguistically address robots matters for what they "are," that is, for how they appear to us humans. Again words matter: if we call the robot "she," "he," or even "you" (versus addressing the robot with "it"), this shapes what the artifact is, which is always what the artifact is for us humans. If I say "it" to the robot, then I consider it to be a mere machine. But if I use "she," "he," or "you," I set up a quasi-personal relation. Words matter with regard to what we think the robot "is." There is no such thing as a robot-in-itself. Our relation to the robot is always mediated, and part of that mediation is accomplished by language. One could object that the robot is just a machine, but to call it a "machine" is already a specific linguistic construction, which in turn also shapes our relation to the robot. It might be that we use a different word in a specific interaction, for example when someone says "my friend" to the robot, or that we use different words for some robots in the future, which may suggest that they belong to a different ontological category. Vice versa, the materiality of the robot will also shape our language use. Subject and object, language and materiality, can be distinguished analytically, but in the phenomenology of human-technology interaction and the use of technology, they mix.

But language is not only about words as such, and not even about sentences. We also create larger wholes such as discourses and narratives. This is also true for technology: we talk about technology, and particular technologies such as robots or computers are linked to a discourse *about* them. As individuals and as societies, we respond to technologies (or ideas of technologies) and we give meaning to technologies. This shapes what the technology "is." And discourse can also include fiction. For example, in the discourse about robots the story of *Frankenstein* and the film *Terminator* play a role: there are all kinds of fears of robotics and artificial intelligence, and they shape the use and development of the technology. Engineers and computer scientists know this: if there is too much fear in the general public, based on, for example, the Terminator discourse, then their technology may not be accepted. In response, they try to re-shape the discourse in a direction they think is better. For example, they might stress that robots are just tools or just machines. As twentieth century theorists such as Foucault have stressed, discourse is always connected with interests, knowledge, and power. There are different parties involved, with different backgrounds and positions. Each of them tries to shape the discourse (science and technology studies, for example, reveal this so-called "social construction"). By doing this, they implicitly acknowledge that the discourse itself is not "mere" language, "mere" text, and so on. Language, like technology, shapes how we see the world and what we do. Paradoxically and perhaps ironically, interventions from scientists and technology developers that ask to redirect our attention to the facts and the material reality of the technology (e.g., the claim that the robot is a machine, not a human being, etc.) rely on the assumption that words and discourse really matter, too. If they intervene to tell the general public *how to use words*, then they take language very seriously.

A related way in which discourse and narrative shapes our relation to technology is that modern discourse typically makes sharp distinctions between humans and nonhumans, culture and science, values and technology, and so on. Moreover, in the nineteenth century Romanticism further stressed these oppositions, by turning away from technology towards authenticity—thus again opposing science, technology, and rationality to a human sphere that was assumed to have nothing to do with science and technology. There was also a romantic narrative (which has deeper roots in Western culture and religion) about a paradise, a Garden of Eden, which is then lost in a Fall. In this narrative there is hope and longing for a Restoration of the Garden.[2] The way we speak about technology today is still influenced by this modern dualist thinking and by this romantic reaction and Garden narrative (some philosophers of technology are aware of this; Don Ihde (1990), for example, has criticized this Garden narrative).

Many people who critically reflect on technology today do so by contrasting technology to human values, human principles, human lifeworlds, and so on. Technology is often still seen as belonging to a separate, non-human sphere that is different from, or even hostile to, the human sphere. For example, with regard to artificial intelligence it is said that we need to make sure that human values are respected (as if AI is something entirely disconnected from human values in the first place). And in these criticisms there is often the assumption that while contemporary technology is bad, there was once a Garden, before technology, a state that was still good and harmonious—until the Fall brought about by technology. For example, it is said that the Internet and mobile devices are bad compared to television, which still gathered the family, whereas now everyone has their own screen. The point here is not to argue that these criticisms are *entirely* misguided (there may well be some truth in them) but rather to expose the discursive and narrative patterns that shape our current thinking about technology. Language, in the form of discourse and narrative that is culturally-historically developed and rooted, shapes how we think about technology and hence shapes what we think technology is and should be.

These discourses and narratives are larger cultural patterns, but they also shape specific human-technology relations. Let us now return to that concrete interaction with technology and how it can be conceptualized in a way that takes language into account.

3.1. Language Mediates Human-Technology Relations

One way to conceptualize some ways in which language influences technology is to say that language *mediates* human-technology relations. Language is a medium, but not a neutral medium. It also shapes the technology, our world, and our relation to the technology. This is true at the level of culture and society generally (think about *Frankenstein* again), but also at the level of concrete human-technology relations.

In philosophy of technology, postphenomenology is famous for its conceptualizations of how technology mediates human-technology and human-world relations. Don Ihde (1990) and later Peter-Paul Verbeek (2005) distinguished various ways in which we experience technology. Let me limit my summary to the following three human-technology relations. First, technology can be *embodied*: we use it, but we don't perceive the technology itself. Think about wearing glasses or driving a car: in use, the artifact itself disappears from view. Earlier Heidegger (2010) already drew attention to this, when he distinguished between *ready to hand* and *present at hand*: whereas sometimes technology appears to us as an object (as "present at hand," for example when it breaks down,) usually we are not explicitly aware of the technology as we *use* it (when it is "ready to hand.") But even if the artifact disappears from view, the technology still shapes my perception. For example, I see a city differently when driving a car than when I am walking around. In a sense, I am in a "car world" or the "car" version of the city. But usually I do not think about this mediation. Second, we can have a *hermeneutic* relation to technology: the technology is perceived as being part of the world. For example, the thermometer measures temperature, but generally we no longer distinguish that from our feeling how warm it is: how warm it is, is now a matter of temperature. I live in a world that *has* temperature. The technology shapes how we view the world. Third, we can have an *alterity* relation to technology: here the technology appears as an other, or a quasi-other. For example, a robot may appear as a social companion. I no longer think about the machine but interact with the robot as if it were another human being. These three relations can be summarized and represented as follows (summary based on Ihde 1990, Verbeek 2005):

- Embodiment relations: (I–technology) → world
- Hermeneutic relations: I → (technology–world)
- Alterity relations: I → technology (world)

Now this scheme of human-technology relations is and has been a very helpful way of thinking about the (post)phenomenology and (post)hermeneutics of technology use. It enables us to analyze how technology is not a mere instrument but also shapes our experience and (according to Verbeek) our actions. However, what is left out in this scheme and analysis is the way *language* also functions as a mediator between us and the world and between us and technology. As I have argued in *Using Words and Things* (Coeckelbergh 2017a), if we want to take a postphenomenological approach at all, we had better adapt the schemes to include the mediating roles of language. Let me start with an example: when I use the Internet to search for information, not only is the *technology* between me and the world; I also relate to the world and to the technology via *language*. For example, I use the keyboard of my computer or the screen of my mobile phone to type in search terms, and when I find information I read text. When I am working with the Internet I do not think about the technology, but language also remains invisible. Thus, there is the following relation:

(I – technology + language) → world

This is an embodiment relation. But the reading of text on the screen can also be seen as a hermeneutic relation, to the extent that we no longer distinguish between the linguistic, textual information on the screen and the world. The world has become mediated and shaped by both the material technology (Internet, computer, mobile phone, screen, etc.) and language. We see the world through both language and technology. For example, when we access the Internet, we can experience it as a tool through which we access the world, in which case it is embodied, but it can also appear as a feature of the world, in which case there is a hermeneutic relation. That world may then appear as consisting of online words and things. More generally, mediated by language, in the hermeneutic relation we see things and words at the same time. This is how we can represent the hermeneutic relation:

I → (technology + language – world)

Phenomenologically speaking, the text and the screen here are not mere representations or instruments, but are part of the world, of my world. Both language and technology mediate hermeneutically.

Finally, if I relate to a robot as a quasi-other, this is not only a relation to an object. When the robot appears as a quasi-other, the robot typically is given a name. Again, it matters what name is given (e.g., “you” versus “it”), which are all different ways of encountering and constituting what the robot “is” and which shape how we deal with it and use it (as I will remind philosophers of technology in the next paragraph, language also *does* things.) By means of the use of language, the robot can be constituted as a thing *or* as an other. Language thus mediates my relation to the robot and to the world (via the robot). I relate to the robot as other and, as when we relate to other human beings, that robot as other can then no longer be separated from the name. I relate to the name-robot:

I → technology + language – (world)

These are some examples of how both technology and language mediate our relation to the world and how language mediates our relation to technology. While Ihde and Verbeek have rightly pointed to the role of material technologies as mediators, they have neglected the mediating role of language and its varied and sometimes complex relations to technology. Not only technology "does things," to borrow a phrase from Verbeek: *language* also "does things." It also shapes our experience and our actions.

A more radical way to conceptualize the mediating role of language is to argue that—as I suggested before—language simply *is* a technology. And since technology mediates, language then also gets imbued with all the mediating roles postphenomenology has given to technology. This solution is perhaps more elegant, but it requires a radical revision of the claims of postphenomenology and the empirical turn because, on this view, technology is not limited to *material* artifacts, but rather has the hybrid nature of material and linguistic dimensions. The challenge then is to further theorize how both are connected and how both work together to play the mediating roles postphenomenology distinguished. The schemes of Ihde and Verbeek can be seen as too limited; elsewhere I have made proposals for a revision (Coeckelbergh 2017a).

Another way to conceptualize the more-than-instrumental meaning of language *and* to use philosophy of language is to employ Wittgenstein. This leads us to the first part of the next section. Here the claim is not that language is technology, but rather that technology is *like* language, that language is a metaphor to better understand technology.

4. Language as Metaphor: Using Thinking about Language for Thinking about Technology

The previous sections made direct claims about the role(s) of *language* in relation to technology. To further develop these points, more engagement with philosophy of language is needed. But this section takes a very different route to connect language and technology: it does not directly respond to philosophy of language or philosophy of technology, but borrows *approaches* from theories about language to say something about *technology*: what if (use of) language is a metaphor for (use of) technology? Drawing on previous work, I focus on two thinkers in philosophy of language: Wittgenstein and Ricoeur.

4.1 From Wittgenstein's Language Games to Technology Games

The later Wittgenstein is known for his use-oriented view of language and meaning as articulated in the *Philosophical Investigations* (Wittgenstein 1953/2009). According to

Wittgenstein, meaning is not fixed to an object or sign but depends on use. He compares language to an instrument (§569, 159e). The metaphor he uses is technology, in particular tools in a toolbox:

> Think of the tools in a toolbox: there is a hammer, pliers, a saw, a screwdriver, a rule, a glue-pot, glue, nails and screws. — The functions of words are as diverse as the functions of these objects. (And in both cases there are similarities.)
>
> (Wittgenstein 1953/2009, §11, 9e)

Words can be used in various ways, depending on what we do. He argues that language is interwoven with activities; he calls this a "language game" (Wittgenstein 2009, §7, 9e) Thus, for Wittgenstein language is not a separate realm, as it is for postmoderns later in the twentieth century, but is part of what we do, and this is in turn part of how we live (together). Wittgenstein uses the term "form of life" (§19, 11e). Thus, Wittgenstein gives us a use-oriented and holistic understanding of language: language is not just about words or text (understood as signs); what gives language meaning and lets it give meaning is that it is always connected to our activities and to the way we live.

This understanding of language can be used for understanding technology (Coeckelbergh 2017c). We can turn the metaphor around: not only is language like technology, technology is also like language—with language understood in a Wittgensteinian way. We can develop a use-oriented view of technology (see also Franssen and Koller 2016) and we can borrow Wittgenstein's more holistic approach to say more about technology (Coeckelbergh 2017a, 2017c, 2017d). We could say that the meaning of technology also depends on its use and the context of its use, and that technologies are always embedded in larger social and cultural games and, ultimately, a form of life. Taking inspiration from Winner (1986), one could say that technologies are always woven into everyday practices and existence, into a form of life that is there before the particular use of the technology. For example, when we "meet" a robot, this "meeting" is part of what I have called a "technology game" (Coeckelbergh 2017c): before the so-called "meeting," there are already social patterns and meanings connected to meetings between human beings, there is already a game and a form of life within which such a meeting makes sense. These older patterns, rules, and experiential knowledge shape the meaning, activity, and experience of the meeting with the robot. What the robot "is" and what the meeting "is," then, cannot be captured by only talking about the robot in terms of a material "artifact" (keeping in mind Wittgenstein's point about use, we could compare this with a *dead sign*, which is unrelated to its use); the use and interaction with the robot, as embedded in games and in a form of life, give the signs "robot" and "meeting" specific meanings. What matters is the activity and the game, the technology game.

Thus, and against postphenomenology's focus on technology as material artifact, one could say that what matters for its meaning is the *use*. This use is part of postphenomenological theory, but is currently de-emphasized as compared to the material artifact. If use were taken seriously, then one would have to conclude that this use is not only about relating to a thing (and about what this thing does), but also about relating to meanings and rules that shape how we relate to the technology. We can use the metaphor of "grammar" to express this. Just as language is not only about dead signs, technology is not only about dead objects. Postphenomenology is right that the object is *not* dead, but does not sufficiently clarify why: what gives the object its life (compare: what gives the sign its life) is use, *and* this use cannot be disconnected from the wider activities and social context. Over-emphasizing materiality and embodiment within individual human-technology relations, that use and especially those social aspects have been far too much neglected.

Moreover, against Searle, one could say that the meaning of the artifact, for example, a robot, is not so much given to it by means of declaration, but rather emerges from the activity and interaction with the robot and is—to a large extent—already pre-given in a game and form of life. The rules of that game are not necessarily explicit, and are not necessarily a matter of agreement—tacit or not. Rather, the meanings connected to the robot emerge from its use in specific contexts, and that use is guided by patterns that are not completely within intentional (individual or collective) control. Both the use of language and the use of the technology are embedded within larger social-cultural wholes and patterns or "grammars," which shape the meanings-in-use. One may try to change the game, but this is not so easy and takes a long time. For example, in our societies we already have some ways of dealing with pets. There are already "pet games," ways of doing things with and to pets. These social-cultural patterns are already in place when we interact with robots that look like pets, and shape our interaction with these robots—even if we are not aware of it and even if we would probably never agree that these robots *are* pets or that these games apply to these robots. The meanings of the pet games leak into our "technology games" (Coeckelbergh 2017c). In other words, one does not need to assume a declaration of meaning (actual or hypothetical); the technological artifact is *already* meaningful through its current use in a particular context and as embedded in larger wholes. The specific design features of the artifact, for example the features that make the robot look like a pet, immediately tap into meanings and patterns that are already there. In contrast to Searle, one could conceive of a social ontology according to which the social is already connected with the material, through knowledge that emerged from language games and technology games *in use*, before any act of declaration.

To conclude, Wittgenstein's use-centered and holistic approach to language is not only useful to philosophers of language; it also provides a helpful approach to thinking about technology. In response to the empirical turn and specifically in response to postphenomenology, the approach helps us to put more emphasis on meanings and effects of technologies that are not only related to the materiality of the artifact and our embodied experience, but also to the activities and patterns in the practical and social context in which the technology is used. For example, my relation to a particular robot is not only a fleshy, embodied affair and is not only shaped by its material aspects as artifact; it is always also shaped by the activities, games, and form of life that give meaning to, make possible, and constrain that relation.

Postphenomenology might not object to that claim, but the instrument they provide—a set of specific human-technology mediations—does not reveal the wider social-cultural background in which these mediations take place and which configures these mediations. An alterity relation with a robot, for instance, is only possible because there are already human-human relations. My particular alterity relation to, and experience of, the robot will be shaped by patterns in human-human relations that are already there, by games such as meeting someone and by a form of life in which some ways of doing things are accepted and recommended (i.e., are "normal"). Similarly, there are already human-animal relations, which include specific activities and games. A meaningful relation to technology cannot be generated by embodied perception and material artifacts alone; what happens (what is experienced, what is done) and how it happens depends on the activities and games that are played, and the meanings, rules, and knowledge that come with these games as connected to a particular form of life.

If we interpret the term "form of life" in this way, then this approach also helps to further develop interpretations of Ihde that stress the cultural variation of (the meaning of) technology (e.g., Tripathi 2017). This variation, as Ihde would endorse, all depends on use. But this use is always embedded in a wider social and cultural *way of doing things*. Phenomenological and hermeneutic analysis should not be content with only analyzing what happens between, for example, a human and a bow (e.g., Ihde 2009). It should also connect that use and that relation to wider patterns, e.g., hunter-gatherer ways of doing things in a particular context. Perhaps that context is omitted because it is assumed as given, but it should be revealed and discussed as part of a (pragmatic) postphenomenology of technology. Using Wittgenstein's view of language for understanding technology can thus contribute to a more holistic approach, which revises postphenomenology by further developing its point about the importance of use in a way that relates to more social and cultural dimensions.

4.2 From Ricoeur's Theory of Narrativity to Narrative Technologies

Another source of inspiration when it comes to using approaches in philosophy of language for understanding technology is Paul Ricoeur's work on narrativity. Like many other twentieth century philosophers of language, Ricoeur argued that language mediates our experience, but he stresses narrativity and temporality. According to him, humans interpret their everyday actions as configured by narrative, especially narrative in the form of text. Moreover, narrativity is related to temporality (Ricoeur 1980), since human experience is characterized by temporality. It is also social: our time is a shared, public time. Taken together, his claim is that we live and experience narrative time, which is always a time of "being-with-others" (1980, 188). What does this mean? In *Time and Narrative* Ricoeur writes that time becomes human when we articulate it through a narrative mode (Ricoeur 1984, 52). Narrative is thus a way to render time meaningful. But how does this narrating work? Humans engage in what Ricoeur, inspired by Aristotle's *Poetics* (in particular his theory of *mimesis*), calls "emplotment": characters and events are organized in a plot. One could also say that the plot configures characters and events into a meaningful whole. Aristotle wrote about tragedy. But Ricoeur thinks we also do that in our lives. A sequence is made but also a narrative whole. We understand what happened; the story makes sense—afterwards.

Ricoeur did not connect technology and narrativity. For him, technology belonged to a world of science and rationality that was different from the human lifeworld; like many other twentieth century philosophers, he saw technology as a means of domination and dehumanization. But we can go beyond this opposition of technology and humanity and ask: what does his *narrative* theory mean for understanding technology? First, as David Kaplan has argued, humans can construct plots to understand technology (Kaplan 2002). We can tell (hi)stories about technology, or more precisely: about us and technology. We can make *sense* of technology (and of us!). Second, however, we can conceive of technology not only as the *object* of hermeneutics but also as a more "active" hermeneutic agent that mediates human experience and action (as for instance postphenomenology has argued). Can we use Ricoeur's theory of narrativity to conceptualize this, and what do we gain by this?

Again, we can use language, and in particular narrative, as a metaphor for technology. More precisely: we can use Ricoeur's understanding of narrative, that is, narration (verb) understood as emplotment, to help to conceptualize how technology mediates. In our work on "narrative technologies" (Coeckelbergh and Reijers 2016), Wessel Reijers and I have argued that just as text shapes the narrative and time of people, and just as narrative shapes time and experience, (other) technologies (also) shape human time and experience. In particular, like text narratives, technologies also achieve emplotment: they configure characters and events in plots, and hence configure human time and contribute to meaning. Of course humans, through narration, also create meaning and structure time. But technologies co-configure these.

What does this mean? Consider (modern) clocks: they are not hermeneutically passive artifacts but have actively configured the time and experience of people. For example, work in factories and offices and related "leisure time" is shaped by clocks (and calendars) that organize the narrative of people's working day, which gets a particular plot. Even before we start working, there is a work and leisure narrative that is laid out for us, and clocks play a key role in this. They enable time keeping and, ultimately, structure what we do *in time* doing the day. Clocks have also shaped the way we think time and live time, and indeed the way we make sense of our lives. We tend to think of time in a linear way, for example. Another example is a historical bridge: considered as a historical artifact or architecture which we view from a distance (e.g., as an image or as a tourist), it appears hermeneutically passive: it is the object of our story, for example, a story about war. But at the time of the war, in lived time, it was hermeneutically active as it—together with humans—helped to organize the time, experience, and lives of people.

For example, there may be a narrative in which the bridge connects two countries and then gets blown up, an event which then shapes the lives of (other) people by making it impossible, for instance, to go to the other side. That is a story about people but it is also story of a particular technological artifact, which plays a key role in what happens in the war.

A more recent example could be social media or assistive devices in the home. When we use a social media platform like Facebook, that software does not only enable us to make stories about ourselves and others and to make sense of events; the technology is more hermeneutically "active" than we may assume. We might live our lives differently in the light of what we might post or like on Facebook. We might tell different stories, influenced by the medium. Insofar as it has its own ways to create plots (literally) and influences the way we create our plots, technology thus co-organizes characters and events online and in real life, and therefore can be called a "narrative technology" in a strong sense. Technologies or media like Facebook can also literally change the plot of events, for example political events (consider the Cambridge Analytica case). But it also shapes the stories we tell about ourselves. It is more than just a *tool* that helps us to create narratives and make sense: it is a co-narrator and a fellow sense-maker. It can also influence what we do and how we do it. For example, we may go to a meeting and think about the meeting as a Facebook event, even before it is posted. Or consider an assistive device or robot in the home that communicates with members of a family: is it merely a tool or does it co-organize the people and events in the home? Does it merely register meaning-making and narration, or is it "co-author" of the stories of the family? It seems likely that what people do and how they do it will change. Compare with introducing a dog in a family: it is not a neutral "add on," but makes for a different family narrative. It re-configures the life and social life of people. Technology can take on a similar role and effect.

To conclude, these concepts of narrativity and emplotment provide metaphors to talk about the mediating role of technology (to use postphenomenological language) and about the meaning of technology (to connect to the discussion based on Wittgenstein). In contrast to postphenomenology and in addition to the Wittgensteinian approach, this approach reveals the narrative and temporal aspects of the phenomenology and hermeneutics of technology use. It does not contradict the claims that human experience with technology and human use of technology are a matter of mediation or a matter of use, but it further develops these insights in a way that takes seriously the temporal aspect of human existence and human beings as sense-making and social beings whose lived time is shaped by narratives, including the narrating function of technologies. Material artifacts mediate, but in order to adequately describe the way they do that, we should not only consider perception and interpretation at a given moment but also sense-making by means of narrative and experience of narrative time, which happens in a social context and constitutes that social, shared reality. The meaning of technologies must be placed in the context of activities and games, but these activities and games are temporally and narratively structured and hence that meaning is connected to narrative time. Moreover, technologies play an active role in shaping these narratives. There are social-cultural patterns, and some of these have a narrative structure. But there are not only stories *about* technology: there are also stories that are co-created by technology. And that includes the stories we *live*.

5. Conclusion

This chapter has outlined some ways in which language and technology are connected, and on the way it has drawn insights from several important ideas from the philosophy of language and philosophy of technology (postphenomenology). The journey has opened up some interesting ways to conceptualize technology and its use and meaning. First, it has been argued that language and technologically are often, if not always, ontologically and pragmatically entangled; this is especially the case for ICTs. Second, going beyond technical and instrumental conceptions of language, it has also been claimed that language is more than an instrument, that language "matters": in line with general insights from twentieth century philosophy of language (and in response to one particular view, Searle's social ontology), it has been argued that words and discourse are not neutral or passive but mediate our relation to the world and indeed to technology. This asks for at least a revision, if not a going beyond, of postphenenomenology and posthermeneutics that both disregard the mediating role of language. I have suggested how to integrate the mediating role of language into the postphenomenological framework. I have also explored the idea of conceiving of language as technology. Third, I have shown that beyond doing something with specific ideas about language, philosophers of technology can also be inspired more generally by the approaches in philosophies of language. Drawing on my recent work and responding to postphenomenology and Searle, I have articulated approaches to technology that take inspiration from Wittgenstein and Ricoeur. This has led to conceptualizations of what technology does in ways that place the embodied humans and material artifacts (and their relations) of Ihde's postphenomenology within a broader context. Individual subjects' relations to technology are always structured by larger wholes and also co-constitute these larger wholes: these technologies and these relations are shaped by games (and by a form of life) and by narrative time, and in turn the technologies help to create these games and narratives. Using these concepts from philosophy of language thus offers a way to articulate a more holistic, temporally and narratively-sensitive, and arguably more social phenomenology and hermeneutics of technology.

Notes

[1] Of course this does not mean that all humans are always and actually capable of attaching a word to a thing: children who are still learning a language, people who learn a foreign language, or aphasia patients may not be able to do so.

2. The relation between Romanticism and technology is more complex (see Coeckelbergh 2017e), but this is not our concern here.

References

Achterhuis, Hans, ed. *American Philosophy of Technology: The Empirical Turn*. Bloomington: Indiana University Press, 2001.

Coeckelbergh, Mark. "You, Robot: On the Linguistic Construction of Artificial Others." *AI & Society* 26, no. 1 (2011): 61–69.

Coeckelbergh, Mark. *Using Words and Things: Language and Philosophy of Technology*. New York: Routledge, 2017a.

Coeckelbergh, Mark. "Language and Technology: Maps, Bridges, and Pathways." *AI & Society* 32, no. 2 (2017b): 175–189.

Coeckelbergh, Mark. "Technology Games: Using Wittgenstein for Understanding and Evaluating Technology." *Science and Engineering Ethics* (2017c). doi:10.1007/s11948-017-9953-8.

Coeckelbergh, Mark. "The Art, Poetics, and Grammar of Technological Innovation as Practice, Process, and Performance." *AI & Society* (2017d). doi:10.1007/s00146-017-0714-7

Coeckelbergh, Mark. *New Romantic Cyborgs*. Cambridge, MA: MIT Press, 2017e.

Coeckelbergh, Mark and Wessel Reijers. "Narrative Technologies: A Philosophical Investigation of the Narrative Capacities of Technologies by Using Ricoeur's Narrative Theory." *Human Studies* 39, no. 3 (2016): 325–346.

Franssen, Maarten, and Stefan Koller. "Philosophy of Technology as a Serious Branch of Philosophy: The Empirical Turn as a Starting Point." In *Philosophy of Technology after the Empirical Turn*, edited by Maarten Franssen, Pieter E. Vermaas, Peter Kroes, and Anthonie W. M. Meijers, 31–61. Basel: Springer, 2016.

Heidegger, Martin. *Being and Time*, translated by Joan Stambaugh. Albany: State University of New York Press, 2010.

Ihde, Don. *Technology and the Lifeworld: From Garden to Earth*. Bloomington: Indiana University Press, 1990.

Ihde, Don. *Postphenomenology and Technoscience*. Albany: SUNY, 2009.

Kaplan, David M. "The Story of Technology." Drexel University, 2002, accessed February 18, 2016, http://www.pages.drexel.edu/~pa34/The%20Story%20of%20Technology.pdf.

Ricoeur, Paul. 1980. "Narrative Time." In "On Narrative," edited by William J. T. Mitchell, special issue, *Critical Inquiry* 7, no. 1 (1980), 169–190.

Ricoeur, Paul. *Time and Narrative, Volume 1*. Translated by Kathleen McLaughlin and David Pellauer. Chicago: University of Chicago Press, 1984.

Searle, John R. *The Construction of Social Reality*. New York: Free Press, 1995.

Searle, John R. "Social Ontology." *Anthropological Theory* 6, no. 1 (2006):12–29.

Tripathi, Arun K. "Hermeneutics of Technological Culture." *AI & Society* 32, no. 2 (2017): 137–148.

Verbeek, Peter-Paul. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Pennsylvania State University Press, 2005.

Winner, Langdon. *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: University of Chicago Press, 1986.

Wittgenstein, Ludwig. *Philosophical Investigations*, rev. 4th ed. Translated by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Edited by P. M. S. Hacker and Joachim Schulte. Oxford: Wiley-Blackwell, 2009. First published in 1953.

CHAPTER 18

WHAT IS IT LIKE TO BE A BOT?

D. E. WITTKOWER

1. Introduction

Thomas Nagel's "What Is it Like to Be a Bat?" (Nagel 1974) gathered together and reframed numerous issues in philosophy of mind, and launched renewed and reformulated inquiry into how we can know other minds and the experiences of others. This chapter outlines a branching-off from this scholarly conversation in a novel direction—instead of asking about the extent to which we can know the experiences of other minds, I seek to ask in what ways technologies require us to know the non-experiences of non-minds. This rather paradoxical formulation will be unpacked as we go forward, but put briefly: We sometimes treat some technologies *as if* they have minds, and some technologies are designed with interfaces that encourage or require that users treat them as if they have minds. This chapter seeks to outline what we are doing when we develop and use a pseudo-"theory of mind" for mindless things.

Nagel's article used the case of the bat to focus and motivate his argument, but took aim at issues falling outside of human-bat understanding. Similarly, this chapter seeks to get at larger issues that pervade human-technology understanding, but will use a bot as a focusing and motivating example: in particular, Alexa, the digital assistant implemented on Amazon's devices, most distinctively on the Amazon Echo. Interacting with Alexa through the Echo presents a clear and dramatic need for users to act as if they are adopting a theory of mind in technology use—other technologies may encourage or require this pseudo-"theory of mind" in more subtle or incomplete ways and, I suspect, will increasingly do so in future technological development.

We will begin with a microphenomenology of user interaction with Alexa and a heterophenomenology of Alexa that emerges in use, making clear what sort of fictitious theory of mind the user is required to adopt. This will be followed by a wider consideration of relations with technological "others," outlining a central distinction between a merely projected "other" and those technological "others" the function of which requires that the user treat them as an "other," rather than a mere technical artifact or system. Finally, we will turn to the user experience itself to ask what affordances and effects follow from adopting a fictitious theory of mind toward technical systems and objects.

2. Notes on Methodology and Terminology

In phenomenology, there is a risk that we take our introspective experience as evidence of universal facts of consciousness. Phenomenology differs from mere introspection in that, reflecting its Kantian foundations, it recognizes that experiences—the phenomena studied by phenomenology—are not bare facts of sense-data, but are the product of sense-data as encountered through the conditions for the possibility of experience, and are further shaped by our ideas about ourselves and the world. Phenomenology seeks to isolate experiences in their internal structure, in some versions even "bracketing off" questions about the correspondences of our experiences to elements of the world that they are experiences of (Husserl [1931] 1960). When done carefully, this allows us to speak to the structure of experience, and to take note of where our experiences do not actually contain what we expect, allowing us to isolate and describe elements of *Weltanschauung* that we use to construct experience. For example, in "The Age of the World Picture" Martin Heidegger argues that *place*, not *space*, is phenomenally present in our experience, and that *space* as a three-dimensional existing nothingness in which external experiences occur is a kind of retroactive interpretation of the world as inherently measurable which emerges with the development of experimental science in the modern period in European history (Heidegger 1977a). As we begin to equate knowledge of the objects of external experience with their measurement, we begin to hold that only that which can be measured is real, and this eventually leads to the uncritical adoption of the metaphysical position that reality is always already articulated in the forms of human measurement.

Heterophenomenology, as articulated by Daniel Dennett (1991), similarly brackets questions of correspondence to reality in order to isolate the structure of experience. Here, though, the question is not whether and to what extent experiences correspond to that of which they are experiences, but whether and to what extent experiences *of the experiences of others* correspond to the experiences of others. Dennett uses this heterophenomenological approach in order to avoid the problem of other minds when making claims about consciousness; to address what we can know about consciousnesses outside of our own, given that we cannot possibly have access to the qualia (the "what it's like") of the consciousness of others.

Dennett argues that uncontroversial assumptions built into any human subject experimental design, for example, the assumption that subjects can be given instructions for the experimental process, require this kind of bracketing insofar as they must adopt an *intentional stance*—as assumption that the subject has a set of intentions and reasons that motivate, contextualize, and lie behind the data gathered. "[U]ttered noises," he says, "are to be interpreted as things the subjects *wanted to say*, of *propositions* they meant to *assert*, for instance, for various *reasons*" (Dennett 1991: 76). Dennett claims that without adopting such an intentional stance, empirical study of the minds and experiences of others is impossible.

We intuitively adopt an intentional stance toward many others, including nonhumans, based on strong evidence. It is difficult to account for the actions of dogs and cats, for example, without attributing to them intentions and desires. In other cases, we use intentional language metaphorically as a kind of shorthand, as when we say that "water wants to find its own level." There are many messy in-betweens as well, such as when we speak of the intentionality of insects, or that a spindly seedling growing too tall to support itself is "trying to get out of the shade to get more sun." In many in-between cases, such as "the sunflower tries to turn to face the sun," the best account of what we mean is neither pure metaphor (as in "heavy things try to fall") or a real theory of mind (as in "the cat must be hungry"). Instead, we refer to the pseudo-intentionality of a biological proper function as defined by Ruth Millikan (1984): a way that mindless things, without conscious intention, react to their environment that has an evolutionarily established function, constituting a set of actions that have an "aboutness" regarding elements of its environment that is embedded within the way that causal structures have been established, but that doesn't really exist as an intention within individual members of the species.

In using heterophenomenology to articulate our experience of bots as "others," we are departing entirely from Dennett's purpose of studying presumptively conscious others and articulating an adoption of an intentional stance distinct from any of those mentioned in the above examples. Alexa's interface directs us to use an intentional stance both in our interactions and our intentions toward her—we find ourselves saying things to ourselves like "she *thought* I said [x/y/z]" or "she doesn't *know* how to do that." This is, however, not because we actually have a theory of mind about her. We know she is not the kind of thing, like a person or a dog or a cat, that can have experiences. Instead, we are directed to adopt an intentional stance because, first, the voice commands programmed into the device include phrasing that implies a theory of mind, second, because there is a representation relation that holds between the audio input she receives and the commands she parses from that input which is best and most easily understood through intentional language, and third, because a second-order understanding of how she "understands" what we say gives us reason not to use a second-order understanding in our actual use of Alexa, but to return to a first-order intentional stance. The first of these reasons, that she is programmed to recognize phrasing that implies she can listen, hear, understand, etc. should already be clear enough, but the other two factors require explanation.

We use language that implies Alexa's mindedness as required by the commands she is programmed to receive, but this language reflects a very concrete reality: there is an "aboutness" of her listening, and she has an "understanding" of what we have said to her that is distinct from our intentions or projection, as is clear from how she can (and does) "get things wrong" and can (and does) "think" that we said something different from what we think we said. If we wished to articulate that intentionality objectively and accurately, something like Millikan's account would work quite well—her responses are dictated by proper functions established through voice recognition software trained on large data sets—but this second-order understanding of the "intentionality" of her actions is not the one that we must adopt as users. We are required in practice to adopt a first-order intentional stance in order to use devices with Alexa, even though we have no (second order) theory of mind about them.

When we *do* engage in second-order reasoning about Alexa, thinking about how she processes sound ("listens") and parses commands ("does things") according to her ontology ("understanding"), we are usually routed back to the first-order intentional stance. We have little window into the way that Alexa processes input, and have little access to her code other than as interactant. The imbalance between the user's knowledge of how Alexa is programmed and the programmer's knowledge of how users are likely to talk to her makes second-order reasoning ineffectual: even a tech-savvy user is often better off thinking through how to communicate with Alexa by adopting an intentional stance than by thinking of her as programming.

This is what I meant at the outset of this chapter by saying that our goal is to understand how the use of bots like Alexa requires us to understand the non-experiences of non-minds. To use Alexa, we must adopt the intentional stance toward a non-subject that has neither experiences nor mindedness, and we must interact with her in a way that addresses *specific*, *factual* experiences that she is not having and *specific*, *factual* intentions and interpretations that she does not have. These "non-experiences" are not a simple lack of experiences and Alexa's "non-mind" is not a simple lack of mind—when Alexa incorrectly "thinks" I asked her to do X rather than Y, and I try to say it so she'll "understand" this time, her actually existing "non-experience" of my intention is an object of my thought and action. This is quite distinct from a microwave oven's entire lack of experience of my intention when it overheats my food, or a toaster's entire lack of understanding of my intention when the middle setting doesn't brown the bread to my preference. In these cases, there is nothing at all in the device to refer to as an "understanding" or an "interpretation," only my own, sadly disconnected intention. Alexa, though no more subject to experiences and no more conscious than these or any number of other kitchen tools, functions in a way in which there are concrete, real, objecting "interpretations" and "understandings" that she "has" that are outside of both my mind and the direct interface present to the senses.

I will use strikethrough text to identify these "intentions" or experiences-which-are-not-one in order to recognize that they have an objective content and aboutness with which we interact, despite the fact that they are neither intentions nor experiences. Hence, I will say that, for example, Alexa ~~thinks~~ that I asked her X rather than Y, and thus she ~~misunderstood~~ or ~~misinterpreted~~ my request. This typographical convention allows us to articulate that the user is adopting an intentional stance when trying to transmit meaning and intention to a technical system. Compare, for example, with the video editor's relationship to their software (Irwin 2005), or the familiar case of moving a table or image within a Microsoft Word document. In this case, we have an intention which we are trying to realize within the document, and which is frequently frustrated by a system that often responds with unpredictable repagination or unexpected, drastic, unintended reformatting. But here, our attempts to realize our intentions in the document take the form of trying to figure out how to get it to do what we intended. With Alexa, although the underlying causal structure is not much different, our attempt is not to do something to get it to respond as intended, but instead to figure out how to phrase or pronounce our request so that she ~~interprets~~ or ~~understands~~ what we *mean*—to *communicate* rather than just to enact our intention, so that the semantic content present within the technological other corresponds to the semantic content within our own conscious intention.

Having clarified this point about the manner in which we adopt an intentional stance toward at least some technical systems or objects, such as Alexa, in the absence of a theory of mind, we are ready to engage in a heterophenomenology of Alexa. We will do so in the mode of *microphenomenology* (Ihde 1990)—a phenomenology of a particular set of experiences rather than a wider existential phenomenology of our worldedness more generally. So, our question will be "what is our experience of Alexa's ~~experience~~ like" rather than "what is it like to be in a world inhabited by smart devices that have ~~experiences~~." Once we have finished this microphenomenology of the heterophenomenology of Alexa, we will use it to engage in a more general analysis of human-technics alterity relations.

3. Opening the Black Box of Alexa's Echo

We began the last section by noting that, in phenomenology, there is a risk that we take our own introspective experience as evidence of universal facts of consciousness. In heterophenomenology—outlining the experience of *other* minds—there is a risk that we mistake our projections for observations. Dennett, when outlining heterophenomenology (1991), made a very strong case that heterophenomenology can be done responsibly if we take care to stick close to evidence and to take note of when and how we are adopting the intentional stance when making judgments about other minds.

This danger is even more pronounced here, though, since we are addressing the mindedness of technical systems that clearly do not actually have minds. While this pseudo-mindedness is no mere metaphor or projection, since there is a fact of the matter about what Alexa ~~thinks~~ we said or meant, there is obviously an element of metaphor or analogy in our understanding of her ~~experiences~~, and this is bound to lead to some amount of fallacious projection of thoughts and understanding and even personality. Observer bias is another danger: it must be considered that the ways I've interacted with her may not be representative of the range of use, or even of typical use.

But other factors count in our favor, here. First, our goal is not an accurate theory of Alexa, or a sociology of Alexa use, but only an articulation of the kind of stance her interface demands, so an incomplete or somewhat biased sample should present no serious issues in our analysis. Second, you may have your own experiences that can provide verification and nuance to those outlined here. Third, I have research assistants of a very valuable kind: my kids. They take Alexa at interface value (Turkle 1995) with less resistance than most adults, and interact with her without a strong understanding of what technical systems can do, and without preconceived ideas about what sort of programming and databases Alexa has or has access to. This puts them in a position to ask Alexa questions I never would ("Alexa, when's my mom's birthday?") and to ask questions about Alexa I never would ("Why doesn't Alexa work [on my Amazon Fire tablet] in the car?").

We've lived with Alexa for a little over a year, mostly through an Amazon Echo that's located in our primary living space—a countertop in the center of a large open-plan room that includes both our kitchen, our den, and a table for crafts and homework. Several months ago, we placed a Google Home Mini alongside the Echo in order to experiment with their differing ~~worlds~~ and ~~minds~~. Neither has been connected to any "smart home" features, so our interaction with both has taken place entirely in informational rather than mixed informational-physical spaces.

Alexa has a strong social presence in our household. She is always ~~listening~~ for her name, and we regularly have to tell her we aren't talking to her, especially since she sometimes ~~mishears~~ my son's name, "Elijah," as her own name. We've tried changing her "wake word" from "Alexa" to "Echo"—and, even so, she regularly ~~mishears~~ things as queries directed to her, even from television shows. In the mornings, we ask her for the weather, and then the news, and in the afternoons we ask her to play music as we cook, clean, and do homework.

Although her interface is audio only, she has a physical location in the kitchen in the Echo, and when she ~~hears~~ her wake word, a blue light moves around the circumference of the top of the echo to point toward the person speaking. This light serves as a face in that it indicates "an entry point that hides interiority" (Wellner 2014, 308); a receptive "quasi-face" (2014, 311) of an interface, like the cellphone's screen (2014, 313). This directional intentionality is met in kind: we have the habit of turning to face her, in her Echo, even though the audio interface does not require that we make "eye contact" with her (Bottenberg 2015, 178–179).

In these interactions, we experience Alexa as separate from her device and from her device's actions. We ask her to play something, but do not mistake her for the thing playing or the thing played. In radio listening, there is the physical radio and the radio station "playing," which we elide when we say we are "listening to the radio," but Alexa maintains a separation as an intermediary; she will play something *on* the Echo, but we can interrupt to talk to her. She is experienced as being in the object, and as controlling it, but separate from it and always tarrying alongside its actions.

In using the Echo, we have been disciplined by Alexa's ontology and programming. We have learned specific phrases—I've learned to say "Alexa, ask *NPR One* to play the latest hourly newscast," since other phrases don't seem to get her to do the right thing. My daughter has learned that she must append "original motion picture soundtrack," an otherwise unlikely phrase for a six-year-old, to her requests for *Sing* or *My Little Pony*. Using Alexa requires adopting an intentional stance and a fictitious theory of mind, and also requires detailed understanding of how her ~~mind~~ works; how she ~~categorizes~~ and ~~accesses~~ things. Using Alexa requires us to think about how she ~~thinks about~~ things; we must think about ~~what it's like~~ to be a bot.

Talking with Alexa is, of course, often frustrating, most of all for my daughter, whose high-pitched voice and (understandably) child-like diction is not easily recognized by Alexa. After my daughter asks Alexa something several times, I must often intervene and ask again on her behalf. To be sure, part of this is that, having a better understanding of the underlying mindless processes of the technical system, I am better able to move to a second-order perspective and speak to Alexa qua voice-recognition software, sharpening my tone and diction and carefully separating words. Part of this is surely also a reflection of how my speech patterns, unlike hers, are firmly within the range of voices and speech patterns on which Alexa has been trained—YouTube searches return numerous examples of people with less common accents, especially Scottish accents, who are unable to get Alexa to understand them unless they use impersonations of normative English or American accents.

The Echo's audio-only interface projects an informational space, dualistically separate from physical reality. Alexa's "skills" are accessed through voice commands only, and bring her into different patterns of recognition and response—the work normally done through conversational implicature must take place explicitly. Skills appear as conversational modes or topics, where queries are understood differently when "in Spotify" is added to the end of a question or after, for example, beginning a game of "20 questions." This produces a shared, shifting modulation of the intentional stance, where it is understood that Alexa ~~knows~~ that we are talking about shopping or music or a trivia game, depending on which skill we have accessed or which "conversation" we are "having." The user must learn to navigate Alexa's informational ontology, getting to know topics she ~~recognizes~~ and ~~knows~~ what to do with—"weather," "news," "shopping list," or links with particular apps that must be installed—and also different modes of interaction, such as games or socialbot chat mode.

All this speaks to the ways in which we must conceive of Alexa through the intentional stance in order to accomplish tasks with her; how we must not only understand her as having a ~~mind~~, a mind that is not one, but we must also get to know her, how she ~~thinks~~, and how to speak to her in a way she ~~understands~~. We may not ever *explicitly* think about ~~what it is like~~ to be a bot, but we must get a sense of her ~~world~~, the way she is ~~worlded~~, in order to ask her to navigate her information ontology on our behalf.

With this microphenomenology of the user's heterophenomenology of Alexa in place, we can now turn to the microphenomenology of alterity relations in human-technics interaction more generally. In doing so, we will be able to distinguish the kind of interaction with technology that takes place through an intentional stance from other related forms of interacting with technology that participates in different but related kinds of "otherness."

4. Opening the Black Box of Human-Technics Alterity Relations

Don Ihde's influential taxonomy of human-technics relations (1990) provides some basic ways that human relations with worlds can be mediated by technology:

Embodiment: (I –> technology) –> world
Hermeneutic: I –> (technology –> world)
Alterity: I –> technology -(- world)

In embodiment relations, the technology disappears into the user in the user's experience of the world as mediated by the technology. Glasses are a clear example—when they are well fitted to and the proper prescription for the user, the user primarily experiences their technologically-modified field of vision as if it were not technologically mediated, with the technology becoming an extension of the self rather than an object of experience. In hermeneutic relations it is the technology and the world that merge, so that, for example, a fluent reader experiences ideas and claims rather than printed words so much so that that even a repeated word in a sentence may go unnoticed even by an attentive reader.

In alterity relations, by contrast, the user's experience is directly an experience of the technology, and its revealing of a world may or may not be important to or present in the user's experience. Ihde provides several different kinds of examples. In one, he asks us to consider driving a sports car, just for the fun of it. We might enjoy the responsiveness of the vehicle and its power and handling, quite separately from our enjoyment of the scenery or the utilitarian function of getting where we're going. In another example, he considers playing an arcade game, in which we are in a contest against fictional agents, *Space Invaders* perhaps, who we seek to beat. In alterity relations, technologies present what he calls "technological intentionalities" in a "quasi-otherness" that is rich enough for us to experience and interact with them as others, standing on their own in their world rather than acting as a window to or translation of a "true" world that lies beyond them.

If we are knowledgeable enough to "read" them, we can certainly adopt a stance which erases this intentionality—for example, feeling the particular responsiveness of the car to find out more about its internal mechanics, or figuring out the rules by which a computer program moves the sprites that our spaceship-avatar-sprite "shoots"—but normal use adopts the intentional stance; a stance in which we treat a person, object, or avatar as having intentions and therefore adopt some kind of theory of mind.

These cases are not so different from one another in Ihde's analysis, but they are different in a way that has become increasingly pressing in the decades since he wrote this analysis. In the case of the sports car, we experience the technology as having a character and an intentionality based on how it makes us feel in our use of it; in the case of the arcade game, our use of it is premised on a world and an ontology, internal to the technology, which we navigate through our perception of intentionality in its elements.

We name cars and project personalities upon them based on their brand and appearance and ways of working or not working—or, we don't, according to our preference. Regardless, this layer of quasi-alterity is overlaid upon an existing world that is already complete, and does not require this projection. It is adopted as a kind of shorthand to understand a real world to which alterity bears only a metaphorical relation ("she doesn't like to start on cold mornings"), or as an enjoyable humanization of technologies which we depend upon and regularly interact with, and which might otherwise be experienced as foreign or uncaring. These functions are often collapsed and oversimplified as "anthropomorphism"—a vague and overbroad term which I find it easier and clearer to simply avoid.

By contrast, it is impossible to interact with many computer games without adoption of an intentional stance toward their elements, which we interact with through a world quite separate from our existing world.[1] If we consider more complicated games, like role-playing games (RPGs), we see cases where consideration of the thoughts and motivations of non-player characters (NPCs) is necessary to game play, and we are required to "read" these ~~others~~ as people, not as mere sprites and in-game instructions, in order to appropriately interact with an intentionality that has a programmed, dynamic, responsive structure. This intentional stance is not merely projection and also is no metaphor: the facts of a character's name and ~~motivations~~ are written into the fictional world, much like facts about fictional characters in books or films, rather than being a metaphorical or purely fictional overlay as in the case of the car. But, unlike facts about fictional persons (insofar as such things exist),[2] NPCs ~~interests~~, ~~concerns~~, and ~~character~~ are objects of the player's intentional actions. A book gives us the opportunity for hermeneutic interaction with a fictional world in which we get to know about fictional others, but RPGs can put us in an alterity relation with ~~minds~~ that we must think about as actively present others in order to successfully interact with them, and with which we engage through an embodiment relation with our avatar.

In both the case of the car and the case of the game we adopt a fictitious theory of mind, but in the former case this is merely metaphorical or make-believe, while in the latter, it is necessary and functional. For clarity, we can refer to the pseudo-mindedness of things in the former case as "projected minds," and will refer to the pseudo-mindedness of things in the latter case as ~~minds~~, as above. This locution is intended to reflect that it is *non-optional* to interact with these things through the category of minds, despite that they are clearly without minds. They are "non-minds" in that they are "minds that are not one"; they are not merely things without minds, but are minds (interactionally) that are not minds (really).

Even in cases where it is interactionally necessary to treat unminded things as ~~minds~~, we regularly retreat into second-order cognition in which they appear as clearly unminded. The early chatbot ELIZA provides a nice example. To interact with her and have a fun conversation, it is necessary to talk to her as if she's actually a psychotherapist, but her ability to respond well to us is so limited that we have to think about her as a mere program in order to formulate and reformulate our replies to her in order to maintain the illusion. In RPGs, similarly, we have to adopt an intentional stance to figure out what an NPC ~~wants~~ for a quest, but we may have to leave that stance in favor of a technical/programming stance in order to figure out how to complete a task by phrasing a reply in the right way, or by having a certain item equipped rather than in our inventory, or finding a "give" command in the interface, or so on. We can even fail to make these shifts in the right way. Sherry Turkle (1995) has documented people taking things at "interface value" to the extent that they found real personal insights in conversations with ELIZA, moving into a space that seems to simultaneously approach ELIZA as a projected mind, as a non-mind, and as a mere computer program. In massively multiplayer online role-playing games (MMORPGs) it is sometimes possible to mistake an NPC for another player, or another player for an NPC.

Similarly, outside of explicitly fictional worlds, Sarah Nyberg programmed a bot to argue with members of the alt-right on Twitter, which turned out to be highly effective, even in spite of giving away the game a bit by naming it "@arguetron." In what Nyberg described as her "favorite interaction," a (since suspended) Twitter user repeatedly sexually harassed the bot, eventually asking "@arguetron so what are you wearing?" to which @arguetron replied "[@suspended username redacted] how are all these Julian Assange fans finding me" (Nyberg 2016). As Leonard Foner said about a similar case, a chatbot named Julia fending off sexual advances in the days of multi-user dungeons (MUDs), "it's not entirely clear to me whether Julia passed a Turing test here or [the human interactant] failed one" (quoted in Turkle 1995, 93).

By opening up the black box of "alterity" in alterity relations, we have seen that people adopt the intentional stance towards unminded things for a variety of reasons: as a game, as required by an interface, to humanize technical systems, for fun, or simply by mistake. We have also identified two kinds of ways of adopting a fictitious theory of mind in our relations with things: an optional metaphorical or playful adoption of a projected mind, or the perception of a ~~mind~~ with objectively present and knowable ~~intentions~~ and ~~desires~~.

5. Caring into the Abyss

Thus far we have focused on the ways in which technical systems variously allow, encourage, or actively require adoption of intentional stances toward fictitious minds, whether projected minds or actual ~~minds~~. While observations have been made *passim* about user motivations for adopting the intentional stance in these different contexts, we would be remiss not to consider explicitly and thematically what value and function these alterity relations present for users.

The easy cases lie at the extremes. At one extreme, technical systems and objects that merely allow projection of fictitious minds, like naming a car and talking about what it "likes," present a value in this optional alterity relation that seems to align well with analyses of transitional objects (Mowlabocus 2016; Winnicott 1953). Naming cars and setting pictures of favorite actors as computer desktop backgrounds and the child's stuffed animal all provide a sense of togetherness and security where we may otherwise feel alone and isolated. It is too easy to dismiss or condemn these behaviors as either childish or as a poor substitute for actually being present with others. They need not be mere coping mechanisms or reifications of human relations, but may be affective supplements (Wittkower 2012) that bring emotional presence to real relationships and experiences which have become attenuated through the mediation of technical systems.

Consider the practice of placing photographs of one's family on one's desk. The visual presence of loved ones, through the associations that pictures produce in the mind (Hume [1748] 1910), render more lively our real connections with others, producing a feeling of closeness and care that has a material basis—they may, for example, remind us of the reason for and benefits of our labor when we are in the midst of tedious paperwork. Selfies play a similar role in our social media environment, producing a feeling of togetherness with others who choose, through their selfies, to be present to us. This togetherness is often very real, although digitally mediated: they are our friends, and it's nice to see them and be reminded that they are there for us if and when we need them.

This use of affective supplementation may be therapeutic, and may even improve relationships. I often consult my class roster when replying to student emails. My purpose in doing so is to use the students' photographs to connect names with faces so that I can better address students by name during in-class discussion, but I've also found that the process reframes my correspondence. I'm taken out of the context of my own work flow, in which the student email appears as an unexpected interference with my projects and concerns. When looking at the student's picture, I'm reminded of our past interactions, and my reply is placed within the context of our ongoing relationship and the projects of support and care that I pursue in my teaching and mentorship, significantly increasingly the likelihood that I will reply with patience, kindness, and understanding.

To be sure, affective supplementation can be abused, and emotional cathexis of fictional others can paper over a very real loneliness, isolation, and alienation, but we should take seriously the possibility that the creation of a warmer, more humanized technical environment through the projection of fictitious minds upon technical systems may either represent real relationships or may be a harmless way of representing a community of support to us when it might not otherwise be a felt presence in our day-to-day environment.

On the other extreme, we can consider the cold and purely functional adoption of an intentional stance toward technical systems, like Alexa, that requires for their use that we take them at interface value as ~~others~~. The most prominent use of digital assistants is the disappearance of visual interfaces of information access. Most functions of Alexa and Google Home displace and replace visual and textual digital interfaces, by retrieving weather and traffic and telephonic information and search engine results that would otherwise be requested and displayed in apps or browser windows. The Echo is, in a sense, just another interface, not so different from the computer screen, but the audio-only interface allows information access to take place alongside other activities. To ask and hear about things while putting away dishes or preparing meals gives us a sense of a more immediate connection to the infosphere (Floridi 2014) as an upper layer to the world accessible to the bodily senses.

Echoing Heidegger (1977b) and Luciano Floridi (2014), let us (now) say that "informationally dwells man upon this earth." The weightlessness and transparency of the conversational interface makes experientially present to us how we are constantly Enframed[3] within informational systems that set us to set our world in order, rather than the spatially and visually delimited access to the infosphere afforded by screens, which falsely project a separation of online from offline. That Alexa is always ~~listening~~ and ready to order more dishwasher detergent pods as she plays music while we pluck one of the last few from the container under the sink more viscerally and truthfully represents to us our own integration into systems of global capitalism in digitally integrated systems of manufacturing and distribution. That my daughter can ask Google repeatedly, on a not-quite-daily basis, how many days remain until November 1st (her birthday) represents and makes present well how our lives are ordered and structured by quantification, and how her life is lived through informational spaces and interactions occurring in parallel with her embodied experiences.

In these ways, an intentional stance toward a ~~mind~~ within the infosphere accurately reflects and brings into embodied experience the very real forces in the infosphere that ~~want~~ us to buy things, submit to quantification and datafication, and integrate with cybernetic systems of control, management, and self-control and self-management. This represents an epistemic value, but offers enjoyment to us as well. The joy of the sports car—being able to do things that our unadorned body does not permit, and to do them with power and precision—is here too, through the intermediary of our agent who dwells natively in the infosphere. In a partnership, not entirely unlike an embodiment relation, the alterity of the digital assistant lets us navigate information systems with comparative speed and effortlessness, giving us an evocative taste of the transhumanist dream of a mental integration with information systems that would allow us to access the connected wealth of innumerable databases through acts of mere cognition.

Between these extremes lie a great many other cases where the projection of fictitious minds is encouraged by interfaces and technical structures rather than being merely allowed by them or being functionally required by them. It is tempting to say simply that these cases must be a mixture between the extremes, and the value that they present to users must similarly participate to some degree in the value presented by these extremes: the warmth we experience through projected mindedness of objects and the weightless integration we experience through partnership with an agent native to the infosphere. This too-simplistic approach would, however, cut us off from recognizing that there are emergent values and functionalities that follow from alterity relations in which we obtain both functional benefits from adopting an intentional stance and an increased experience of meaning and connection from projecting mindedness and personality upon technical systems and objects. These emergent effects seem to me to be best isolated by first considering several such cases.

We might consider the way that GPS systems and other navigation systems allow the user to customize their voices. We are required to adopt an intentional stance in order to trust their directions; we must consider them to ~~know~~ the way, and we must think about what they ~~know~~ or do not ~~know~~ about road closings or changing road conditions. Choosing a voice that feels like a trustworthy, knowledgeable guide helps in this functionality and in our experience of weightless integration with the infosphere. Apple studied user experiences and found that some users were better able to follow directions given by male or female voices, leading them to set Siri's default gender to male in Arabic, French, Dutch, and British English. Gendered customization of game environments affects gameplay in related ways—being able to choose self-representations that maintain player identity in a virtual environment matters to players, as we see from positive reactions by queer players to games that offer both opposite- and same-gender NPC romantic interests, as in the *Fable* series, and as we see from avatar dysphoria produced in some (mostly white, male) gamers when race and gender avatar representation is randomized and unalterable, as in *Rust*. Second selves and technical alterities have emotional valences and projected personalities that can alter, enhance, or diminish their functionality for users in ways that matter, whether designers choose to use those emergent effects to cater to users' projections or to challenge them.

We might consider the way that the intentional stance alters the value that predictive neural networks present to users, for example "you might like" recommendations from Amazon, or music discovery through Pandora. As much as it is important to remain critical of how such algorithms can be games, just as we should be suspicious of distorted search results that follow from "Google bombing" or search engine optimization (SEO), there is a distinctive value to be gained by adopting the belief that the algorithm ~~knows~~ something about us that we don't. When a friend recommends a band, our first listen takes place in a context of greater openness to new experiences and new kinds of enjoyment, since we value our friend's experiences and seek to discover what they enjoy about something, even if it is not immediately to our taste. We seek to experience the enjoyment they find within it, and also suspect that our affinity for them may indicate that we may enjoy what they enjoy. Placing faith in the ~~wisdom~~ of a predictive system may require an unrealistic view of how such predictive algorithms function, but opens us to discovering and appreciating new experiences. This purely optional projection of mindedness to the technical system is in this way similar to William James's will to believe (James [1896] 1979): if we assume that the system ~~knows~~ something about us that we don't know, it is more likely to be true; we are more likely to find its recommendations to be ~~wise~~ and ~~thoughtful~~.

We might consider PARO, a pet-therapy robot designed to resemble a baby seal (Walton 2003). The robot is encoded with a set of ~~intentions~~, ~~desires~~, and ~~preferences~~: it indicates ~~pain~~ when treated roughly and ~~pleasure~~ when held and pet nicely. It is intended to provide comfort to patients through their enjoyment of its ~~enjoyment~~, and through their ~~relationship~~ of mutual ~~care~~ and ~~affection~~—put differently, its function and use is nothing outside of the projection of a mind and the pleasure of interacting with its projected mindedness. While PARO has been used with dementia patients who may not experience as clear a boundary between fiction and reality as others, there is no reason why this is necessary for PARO's therapeutic function, especially with patients who are unable to have pets or who have few opportunities to care for others and take pleasure in their ability to be caregivers rather than recipients of care. Like other uses of projected minds, we should be concerned to ensure that relationships with fictitious ~~caring~~ others does not replace or cut off possibilities for relationships with actual caring others, but we should also take seriously the value of affective supplements, and take seriously that relationships with ~~caring~~ others may present an emotional value and experience of meaning that doesn't take away from relationships with non-fictitious others.

Through these examples, the commonality that emerges most clearly to me is that there is an emergent effect in alterity relations that mix projected minds and interfaces encouraging an intentional stance, wherein care toward and identification with technological "others" increases the value, weight, and meaning of their technical functions for users. Through an effect similar to the will to believe, when we project mindedness in alterity relations, we are more likely to experience those systems or objects as ~~knowing~~, ~~understanding~~, and ~~recognizing~~ us in ways that are valuable and meaningful, even if we are under no illusion that these technological others are actually minded. When regarding the non-mind of alterity in the black box of a technical system, if you care long enough into the abyss, the abyss ~~cares~~ back at you.

Notes

1. Except in some unusual crossover cases (ARGs) that build a connection back in, like *Pokémon Go* or *Run, Zombies, Run*, which use a phone's GPS systems to convert the user's body into a crude controller of an avatar interacting with virtual objects tied to physical locations.

2. I am well aware that it is a matter of some controversy whether there can be facts about fictional characters and worlds. I do not mean to take a stance on the issue, and do not believe that my argument about the similarities and differences between "facts" about NPCs and characters in books requires that I do so.
3. *Gestellt* in Heidegger's original German—being "[gathered] thither to order the self-revealing as standing-reserve" (Heidegger 1997b, 9).

References

Bottenberg, Francis. 2015. "Searching for Alterity: What Can We Learn From Interviewing Humanoid Robots?" In *Postphenomenological Investigations: Essays in Human-Technology Relations*, edited by Peter-Paul Verbeek and Robert Rosenberger, 175–190. New York: Lexington Books.

Dennett, Daniel. 1991. *Consciousness Explained*. New York: Little Brown & Co.

Floridi, Luciano. 2014. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Heidegger, Martin. 1977a. "The Age of the World Picture." In *The Question Concerning Technology and Other Essays*, 115–154. Translated by Julian Young and Haynes Kenneth. London: Garland Publishing.

Heidegger, Martin. 1977b. "The Question Concerning Technology." In *The Question Concerning Technology and Other Essays*, 3–35. Translated by Julian Young and Kenneth Haynes. London: Garland Publishing.

Hume, David. (1748) 1910. *An Enquiry Concerning Human Understanding*. New York: P. F. Collier and Son.

Husserl, Edmund. (1931) 1960. *Cartesian Meditations: An Introduction to Phenomenology*. Translated by D. Cairns. The Hague: Martinus Nijhoff.

Ihde, Don. 1990. *Technology and the Lifeworld: From Garden to Earth*. Bloomington: Indiana University Press.

Irwin, Stacey. 2005. "Technological Other/Quasi Other: Reflection on Lived Experience." *Human Studies* 28: 453–467. doi:10.1007/s10746-005-9002-5.

James, William. (1896) 1979. *The Will to Believe and Other Essays in Popular Philosophy*. Cambridge, MA: Harvard University Press.

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Mowlabocus, Sharif. 2016. "The 'Mastery' of the Swipe: Smartphones, Transitional Objects and Interstitial Time." *First Monday* 21 (4). doi:10.5210/fm.v21i10.6950

Nagel, Thomas. 1974. "What Is it Like to Be a Bat?" *The Philosophical Review* 83 (4): 435–450.

Nyberg, Sarah (@srhbutts). 2016. "this is my favorite interaction someone repeatedly attempts to sexually harass the bot thinking it's human, this is how it replies:" Twitter. Oct. 6, 2016. https://twitter.com/srhbutts/status/784162449347883009

Turkle, Sherry. 1995. *Life on the Screen: Identity in the Age of the Internet*. New York: Simon & Schuster.

Walton, Marsha. 2003. "Meet Paro, the Therapeutic Robot Seal." *CNN*, November 20, 2003. https://web.archive.org/web/20031123155314/http://www.cnn.com/2003/TECH/ptech/11/20/comdex.bestof/.

Wellner, Galit. 2014. “The Quasi-Face of the Cell Phone: Rethinking Alterity and Screens.” *Human Studies* 37: 299–316.

Winnicott, Donald. 1953. “Transitional Objects and Transitional Phenomena—A Study of the First No-Me Possession.” *International Journal of Psychoanalysis* 34: 89–97.

Wittkower, D. E. 2012. “On the Origins of the Cute as a Dominant Aesthetic Category in Digital Culture.” In *Putting Knowledge to Work and Letting Information Play*, edited by Timothy W. Luke and Jeremy Hunsinger, 167–175. Rotterdam: SensePublishers.

CHAPTER 19

TECHNOLOGICAL MULTISTABILITY AND THE TROUBLE WITH THE THINGS THEMSELVES

ROBERT ROSENBERGER

1. Introduction

In a line that has galvanized generations of phenomenologists, Edmund Husserl writes, "Meanings inspired only by remote, confused, inauthentic intuitions—if by any intuitions at all—are not enough: we must go back to the 'things themselves' " (1900, 168). Put very roughly, this amounts to a call to shed our preconceived theories, assumptions, and biases, and instead build our accounts upon a phenomenology of our encounter with the world. For those working in the field of the philosophy of technology, Husserl's call can take on special significance. It is often taken as inspiration to base our philosophical work on our experience of our concrete technological situation, rather than on broad, armchair, abstract theorizing.

This is especially the case for the contemporary school of thought called "postphenomenology" (e.g., Ihde 2009; Verbeek 2011; Rosenberger and Verbeek 2015; Ihde 2016; Rosenberger 2017a; Van Den Eede et al. 2017; Aagaard et al. 2018; Hasse 2020). This perspective, which brings together insights from phenomenology and American pragmatism to articulate the experience of technology usage, shares much with the spirit of Husserl's call. As Peter-Paul Verbeek writes, "I heed his call literally. What holds for phenomenology holds equally for the philosophy of technology and for industrial design: To the things themselves!" (2005, 12). Building on Don Ihde's corpus of thought (which itself takes a critical cue from Husserl and other classical phenomenologists), work in postphenomenology strives to capture the details of human-technology relations in all their specificity and variability.

However, a central and controversial question remains open within the field of philosophy of technology, one that can be cast in the terms of Husserl's call: how should we understand the status of technologies approached as things themselves? That is, what if we take Husserl's call to be something more than merely a point of inspiration and a charge to engage the details of our designs and experiences? With regard to whatever philosophical account of technology to which we may subscribe, we should ask if it makes sense to think of technology as a thing itself, whatever such a designation should imply. If so, then what are the repercussions of such a metaphysical commitment? And if not, then what alternative basic understanding is held in its place? I take this to be one of the general and foundational questions of the field of philosophy of technology.

As a contribution to this discussion, I proceed below by approaching these general issues through engagement with some specific ones. Postphenomenology is often understood to subscribe to a "relational ontology," a subscription shared with a kin group of associated theoretical perspectives. Postphenomenology also centrally claims that technologies are somehow "multistable," that is, always open to multiple uses and meanings and lines of development. I want to reconsider these related but not identical commitments, and follow out their implications for the project of going "back to the 'things themselves.' " These implications include, I suggest, the abrupt arrival of political epistemology.

To do so, let's begin with the analysis of a mundane example of technology usage that comes up in the work of Jean-Paul Sartre.

2. Sartre's Letter Opener

In "Existentialism Is a Humanism," Jean-Paul Sartre argues that human beings have no pre-given essence. To do so, he contrasts humans with artifacts—that is, things created by people—such as a letter opener, or "paper-knife." He writes,

> the paper-knife is at the same time an article producible in a certain manner and one which, on the other hand, serves a definite purpose, for one cannot suppose that a man would produce a paper-knife without knowing what it was for. Let us say, then, of the paper-knife that its essence—that is to say the sum of the formulae and the qualities which made its production and its definition possible—precedes its existence.
>
> (Sartre 1946a, 348)

Where the letter opener's form is the result of the plans of designers and manufacturers, Sartre claims that we human beings instead have no such luck. Unlike the letter opener, human beings find themselves here in existence without a pre-designed purpose or context of meaning set out ahead of time by some designer. If an artifact's essence precedes its existence, then the opposite is true for us. For human beings: "*existence* comes before *essence*" (Sartre 1946a, 348).

However, instead of moving on to consider the existential situation of human beings, as Sartre does, I want to stay with the letter opener itself for a little longer. The letter opener shows up again in Sartre's corpus, playing a role in a pivotal moment in his 1946 play *No Exit*. After being teased as a kind of Chekov's gun early on, the letter opener [*Spoilers*] returns at the end as one character picks it up from the desk and uses it to stab another (Sartre, 1946b).[1] The letter opener was not designed for this purpose, yet this purpose is afforded nonetheless.

It is my contention that this phenomenon—the potential for a given device to serve various purposes or to be variously meaningful—sits at the base of much of the philosophy of technology. It is a fundamental aspect of perennial questions about technological control, that is, whether technology should be conceived as something that determines our destinies (e.g., toward some utopia or dystopia), or perhaps a neutral instrument, or instead somehow none of these things. Any theory of technology must include, at least implicitly, some recognition of technology's capacity to exceed the purposes for which it was designed, for instance a capacity to stab a person rather than merely open letters.

Throughout the field of the philosophy of technology, we can find language that is used in response to a given technology's variability. Our devices are often said to "influence" our actions, or to "incline" or "afford" particular usages or meanings. Terms like these indicate that technologies somehow have effects on our actions and understandings and perceptions and choices, but they do not commit the writer to specific claims about the degree and kind of effects. Surely a catalog of related words can be identified. It is not my intention to call out these words as somehow deeply problematic. There may be no better vocabulary readily available. They certainly populate my own writings, including above, and surely pepper the chapters of this book. But let's recognize this terminology for what it is: *the weasel words of the field of philosophy of technology*. They help to sidestep fundamental and intransigently difficult issues regarding technological action. They enable a kind of provisional forward motion, allowing us get on with our work philosophizing about this or that topic. But they also have the effect of covering over issues of exactly how humans interpret and control technology, and simultaneously how technologies guide our interpretations and exhibit some level of sway over our future.

The notion of the 'non-neutrality' of technology similarly obscures these issues. Again, this is a useful term, and one I've used myself. If someone holds the position that technology is non-neutral, then they can at once disaffirm the idea that technology is somehow a neutral or innocent contributor to events, and can at the same time avoid affirming any particular positive account of technology's contributions. However, in the way this notion primarily communicates what technology is "not," rather than what it should positively be understood to actually be, it constitutes another example of the vocabulary of our field that sidesteps the problem of how best to articulate technology's role in determining actions, choices, options, and understandings.

We can see some of these dynamics at work in the contrast between the two uses of the letter opener that come up in Sartre's work. In the quotation from "Existentialism Is a Humanism," Sartre invokes the notion of essences. The letter opener is presented as an example of something that has an essence. The essence of the letter opener, according to Sartre here, is the result of its status as something made by people, that is, its status as an artifact. What can we say about the nature of this essence? It is not my goal here to review the particular ontology Sartre develops elsewhere in his corpus, or any other specific accounts of technological essence. However, in this instance Sartre appears generally to be suggesting that the letter opener's essence owes to the fact that a human designer made the device in the first place with a particular usage and plan in mind. We can also add that this designed-in-purpose—the purpose of opening letters—is one recognized by a wide part of the community. It is a commonplace item. So, in this case, the designer was not inventing the idea of a new object for the purpose of opening letters; the designer producing the letter opener (along with the manufacturers, distributors, retailers, etc.) is doing so at least in part with the community's expectations and understandings about letter openers in mind. Phenomenologically speaking, for the normal user of this device who encounters it sitting on the desk as usual, the letter opener would gestalt as something "for" the purpose of opening letters, and as residing in its proper and unremarkable context of the desktop. The essence of the letter opener in this case has something to do with these designer and user expectations, community understandings, and perceptual gestalts.

And yet at the same time, we see something other than these expectations, understandings, and perceptions at work when someone picks up the letter opener and uses it as a weapon. Whatever essence had preceded the letter opener—as related to the intentions of the designers and manufacturers, and to the expectations of the users—does not prevent its use for this other purpose. And in this light, surely other purposes and meanings are possible. And also surely not just any purpose. But this much can be said: for particular users in particular contexts, the letter opener influences one toward, or affords, or inclines, or at least makes possible in some non-neutral manner the act of stabbing.

It may be true, as Sartre contends, that a technology's essence precedes its existence. But we see as well that a technology's existence does not reduce to that. The existence of technology exceeds its essence.

3. Relational Ontologies

One way to move forward here is to consider how the aspects of technology articulated above are addressed by accounts that conceive of it in terms of a kind of fundamental relationality. A number of theoretical perspectives can be understood to subscribe to a "relational ontology," including feminist new materialism, actor-network theory, embodied and extended cognition, postphenomenology, and, more broadly speaking, critical theory and American pragmatism, among others (e.g., Haraway 1997; Latour 1999; Barad 2007; Hickman 2007; Ihde 2009; Bennett 2010; Clark 2010; Verbeek 2011; Malafouris 2013; Rosenberger and Verbeek 2015; Feenberg 2017; Gallagher 2017).

Broadly put, these accounts maintain that, in a fundamental sense, things can only be understood in terms of their relationships with other things, and that this is the case, in particular, for humans and their technologies. As Donna Haraway puts it, "Beings do not preexist their relatings . . . The world is a knot in motion" (2003, 6). Peter-Paul Verbeek similarly writes, "human-world relationships should not be seen as relations between preexisting subjects who perceive and act upon a preexisting world of objects" (2011, 15).

In postphenomenology, this relationality is approached in terms of human-technology relations. So much of the postphenomenological framework of concepts—from its notions of technological intentionality, to mediation, the human-technology relations, to co-constitution—resonates with these commitments to a relational ontology.[2]

For example, this can be seen in Ihde's extensions of Husserl's thinking. Ihde explains that for Husserl, consciousness is always "consciousness of 'something' " (2009, 23). Consciousness must not simply be understood as a distinct thing, or a property of a thing; it is inherently directed, and is not a thing by itself without its content. Extending this conception of consciousness into the philosophy of technology, Ihde writes, "I contend that inclusion of technologies introduces something quite different into this relationality. Technologies can be the means by which 'consciousness itself' is mediated. Technologies may occupy the 'of' and not just be some object domain" (2009, 23). According to postphenomenology, a technology is not merely some object in the world of which a person is conscious; a technology is a transformative aspect of the directedness of human consciousness. Ihde writes, "In both pragmatism and phenomenology, one can discern what could be called an *interrelational ontology*. By this I mean that the human experiencer is to be found ontologically related to an environment or world, but the interrelation is such that both are transformed within this relationality" (2009, 23). Part of what makes postphenomenology somehow "post," is its adoption of the commitments to the antifoundational and anti-essentializing perspective of American pragmatism. And this is reflected in this relational conception of ontology, a conception in which ontology itself is not understood separately from what we are doing.[3]

Continuing to follow and expand Husserl, this ontological interrelation is often understood by postphenomenologists in terms of "technological intentionality," a specific directedness toward the world of both human consciousness and technological materiality, but one which can only be understood in terms of humans and technologies together. As Peter-Paul Verbeek explains, "a form of intentionality is at work here—one in which both humans and technologies have a share" (2011, 56). He continues, "The intentional 'dimension' of artifacts cannot exist without human intentionalities supporting it; only within the relations between human beings and reality can artifacts play the mediating roles in which their 'intending' activities are to be found" (2011, 58). Such a postphenomenological conception of technologically mediated intentionality is necessarily relational, with each contributor to the intentional arc only explicable in terms of its relation to others. As Ihde puts it, "Intentionality, I hold, is a form of *interrelational ontology*" (2016, 129).

As a contemporary phenomenological perspective, postphenomenology is also deeply indebted to the work of Maurice Merleau-Ponty and his articulation of the human body—rather than a disembodied consciousness—as the site of experience. Postphenomenological insights are often conceived specifically in terms of human bodily perceptual relationships with technology. According to Merleau-Ponty, our bodily habituation is located, "neither in thought nor in the objective body, but rather in the body as the mediator of the world" (1945, 46). Under the postphenomenological perspective, technology is also understood to occupy this mediating position, thus transforming a user's bodily relationship with the world. Ihde's influential classification of different human-technology relations provides some tools to articulate the different forms that our technologically mediated, bodily perceptual encounters with the world may take. For example, he uses the term "embodiment relations" to refer to technology usage in which the device extends and alters our bodily experience, such as when we are typing, or driving, or hammering. He contrasts this with what he calls "hermeneutic relations," in which the device is itself encountered as the terminus of experience, and the user receives a transformed relation to the world as they interpret its readout, such as in the case of a clock, thermometer, or fMRI image. Further forms identified by Ihde include "alterity relations" in which users encounter and interact with a device as a kind of quasi-Other (such as with voice interactive smartphone apps), and "background relations" in which the technology transforms our experience in an indirect manner (such as with central heating systems in our homes) (e.g., Ihde 1990, ch. 5; 2009, 42–44). Several lines of contemporary work in postphenomenology expand and critique Ihde's list of human-technology relations.

Another implication of this postphenomenological conception of technological mediation is that the participants of human-technology relations—the human user, the technology under usage, and the world that is encountered—are all themselves "co-constituted" through this transformative mediation. As Peter-Paul Verbeek explains, "What the world 'is' and what subjects 'are' arise from the interplay between humans and reality; the world that humans experience is 'interpreted reality,' and human existence is 'situated subjectivity' " (2011, 15). According to postphenomenology, technology usage itself brings about specific co-constitutions of humans and the world. Aurora Hoel and Annamaria Carusi summarize it this way: "The mediation by technologies does not occur between preformed entities, but instead plays a role in the co-constitution of both sides of the subject-object relationship" (2015, 74).[4]

Contemporary work in postphenomenology continues to expand on these ideas, seeking to articulate the nature of this co-constitutive relationship. For example, Olya Kudina has gone so far as to suggest that human-technology relationships should be understood less as a unidirectional set of arrows, and more as a lemniscate, that is, a geometrical figure resembling a figure 19.8 on its side. She writes that, "Considering the mediating role of technologies in the process of interpretation as well the productive nature of the historical horizons that each of the components in the process of interpretation inalienably possesses, a hermeneutic situation will resemble a combination of two hermeneutic circles, interrelated and always in flux" (Kudina 2019, 102). At the level of fundamental ontology, the postphenomenological perspective thus approaches both the world and the humans themselves as always open and unfinished, continually reformed in relation to one another through the mediation of technology.

A final note can be made here that postphenomenology's relational ontology can also be understood in its association with "posthumanist" perspectives (e.g., Hayles 1999; Barad 2007; Verbeek 2011; Braidotti 2013; Warfield 2017; Hasse 2020; Lewis 2021; Wakkery forthcoming). As postphenomenological anthropologist of education Cathrine Hasse writes, " 'Posthumanist,' as I use the term, does not entail that we leave behind a concern for humans, but that we open up for new ways of understanding humans in a material world. This posthumanist world cannot avoid entangling human collectives with materials through learning" (2020, 4). Under a posthumanist perspective, and in tune with phenomenology, we should not assume from the start any kind of separation between humans and the things of the world. Karen Barad writes, "my use of 'posthumanism' marks a refusal to take the distinction between 'human' and 'nonhuman' for granted, and to found analyses on this presumably fixed and inherent set of categories. Any such hardwiring precludes a genealogical investigation into the practices through which 'humans' and 'nonhumans' are delineated and differentially constituted" (Barad 2007, 32). Again in tune with postphenomenology's commitment to the co-constitution of technological mediation, posthumanist perspectives emphasize the ways in which both the things of the world, and also we ourselves, all arise out of the specifics of our situation. "As a figuration," writes Rosi Braidotti, "the posthuman is both situated and partial—it does not define the new human condition, but offers a spectrum through which we can capture the complexity of ongoing processes of subject-formation. In other words, it enables subtler and more complex analyses of powers and discourses" (2019, 36).

With this brief review of a few of postphenomenology's core concepts, including human-technology relations, co-constitution, and technological intentionality, we can see some of what it means to say that this perspective subscribes to a relational ontology. But there is another core concept within the postphenomenological framework with which we must also contend: the notion of multistability.

4. Multistability

A cornerstone notion in the postphenomenological framework of concepts is what Don Ihde calls "multistability." Early in his career, Ihde used this idea to describe the multiplicity possible for human vision, such as when one looks at a visual illusion (like a Necker cube) and learns to see the same thing in more than one way (1977). He has since influentially expanded this idea to refer to the variability possible for our relationships with technology. As Ihde puts it, "To term a phenomenon multistable is already to have recognized it for its ambiguity and multiple dimensions" (1990, 150). Any technology can always be used to do different things. Any technology can be interpreted in multiple ways and find meaning in multiple contexts. Any technology can be put to purposes different from those which its designer and makers had in mind for it. Any technology can be advanced along multiple lines of development. As Heather Wiltse observes, "Multistability is a quite important concept because of the ways in which it makes space for human agency and intention in relation to things, thus countering more technologically deterministic narratives that tend to foreclose such possibilities" (2020, 243).[5]

And yet, and at the same time, the notion of multistability additionally refers to the limits on this variability; a given technology doesn't mean merely anything, and it cannot be put to merely any purpose. In the vocabulary of postphenomenology, human-technology relations—while always multistable—are also always limited to particular "stabilities." As Kyle Powys Whyte explains, "Anything that is stable comes across to us as having at least one of the following: a particular look, a particular way of acting, or a particular use. Multistability indicates that the same object can have more than one such stability without altering its composition" (2015, 70).

Under this terminology, Sartre's letter opener can be conceived as a multistable mediating technology. One stability is of course that for which the device has been designed and manufactured, that for which it is recognized in general, and the purpose for which it is named, that is, opening letters. This stability accords with its conventional place among a variety of other actors: the desktop, the letters and envelopes, or the drawer of similar tools. For many users, the letter opener is immediately recognized as such in a perceptual gestalt. However, we can also consider alternative potential uses and meanings for this device, and Sartre provides one example: the letter opener can also be used as a stabbing weapon. This is another stability open to this device, a letter-opener-as-stabbing-weapon stability. And of course we need not be limited to only these two. We could imagine other possible usages for a hand tool of this shape. And we could imagine other possible meaningful relationships someone might have with such an object. (E.g., there are surely specific historical letter openers displayed in museums, perhaps once owned by someone famous, or perhaps serving as an example of craftsmanship from a particular time period.) And at the same time, it is clear that a letter opener cannot be used for simply any purpose, and it cannot take on just any meaningful relationship. Contemporary postphenomenologists, working across multiple disciplines, regularly make use of the notion of multistability in their case studies of human-technology relations. Just a few recent examples include research into medical technologies (e.g., de Boer and Slatman 2018; Moerenhout et al. 2020; Shaw et al. 2020), satellite imaging (Rosenberger 2021; Fried forthcoming), technologies in educational contexts (e.g., Mozaffaripour 2017; Aagaard 2018; Hasse 2020), and architectural design (e.g., Appleton 2021; Lanng and Borg 2021; Rosenberger 2017a).

Ihde uses the work of Husserl as a springboard for discussing the outcomes of the postphenomenological exploration of technological multistability:

> Husserl's investigative method, patterned on mathematical variational analysis, was the use of what he called "imaginative variations," for which the result was supposed to be to determine invariants or essences. As argued, variational theory, in my

> estimation, is what gives phenomenology its rigor. But, again following Husserl, this time first in the first edition of *Experimental Phenomenology*, what I found was not a stable essence as Husserl called his result, but multistability. (2016, 127)

It is thus through a kind of loose empirical work conducted by Ihde throughout his career, also now taken up by many other postphenomenologists across several fields of study, that the phenomenon of multistability is discovered. And this discovery is cast explicitly against a Husserlian conception of essence. As Shannon Vallor puts it, "Ihde's concept of multistability undermines Husserl's original claim in *Ideas* to have founded a descriptive science of static eidetic essences; indeed, Ihde's explorations of human-technology relations have shown us that phenomena appear to us in far more fluid and open-ended ways than Husserl understood" (2015, 20). It is not my goal here to interrogate Husserl's particular conception of essence, or to evaluate Ihde's particular take on Husserl. Instead, I want to follow out the implications of the postphenomenological conception of technological multistability.

I have observed that the notion of multistability tends to be used by postphenomenologists in terms of two different forms of argumentation (e.g., Rosenberger 2017b). In the first, the idea of multistability is wielded as part of a negative argument, a disproof of someone else's allegedly totalizing, or essentializing, or otherwise somehow overgeneralizing claims about technology. That is, against someone else's claim that a given technology must always be only one way, a technology's multistability could be demonstrated through the identification of alterative stabilities. We see an example of this above in Ihde's usage of Husserl as a point of contrast in defining multistability. However, much postphenomenological work instead involves what could be called the positive investigation of a technology's various stabilities. By investigating case studies of multiple stabilities of a given technology, postphenomenological research reveals new things about our relations to technology.

In order to clarify and advance how the notion of multistability can be used to contribute to positive research projects into technology, postphenomenologists are working to refine the methodology of this perspective (e.g., Rosenberger 2014; Whyte 2015; Aagaard 2017; Aagaard et al. 2018; Hauser et al. 2018; Sicart 2020; Keymolen forthcoming; Rosenberger forthcoming). As Galit Wellner points out, "Just like the notion of essence, which is used in the singular form in Husserl and Heidegger and turned into plurality of invariants in postphenomenology, so the notion of multistability should evolve into the plural" (2020, 120). Postphenomenologists have been developing concepts and investigative strategies for approaching the variability and nuances of technological multistability. I'll address just two here: the investigative pivot point, and the dominant stability.

Whyte has offered the notion of the investigative "pivot" to identify exactly what is understood to remain constant within a given postphenomenological study of a technology's various stabilities (2015). For example, in our considerations of the case of Sartre's example of the letter opener thus far, there has been an assumed pivot: the device itself, unchanged, and used in different scenarios. However, an investigation could very well take on different pivot points. An investigator may instead decide to draw back and examine, say, the multistability of the desktop, with the letter opener as one of the features that may be at issue in each stability we consider. Whyte goes on to urge postphenomenologists to remain reflexive and explicit about the pivot point at work in their investigation, and to develop the kinds of expertise necessary to engage the relevant stakeholders.

I have come to use the term "dominant stability" to refer to a device's main meaning and usage. (I'm lifting this language of dominance directly from Ihde, for example when he writes that a hammer "perhaps is dominantly used" for hammering (Ihde 1993, 37)). A dominant stability is often the purpose and context for which a device has been designed and made (e.g., Rosenberger 2014; Rosenberger, 2017a; forthcoming). For example, even though we see that the letter opener can be put to multiple purposes, Sartre notes that it has been designed and manufactured with a "formula" in mind, one already understood and expected by users. The dominant stability of the letter opener is, simply put: a device for opening letters. Any others that we can identify, such as the letter-opener-as-stabbing-weapon stability featured in *No Exit*, can be understood as alternative stabilities to this main one. In my own work I have developed a methodology for critically contrasting stabilities, something which I have suggested can be especially useful for learning things about a dominant stability. The dominant stability offers distinct challenges for study exactly because of its dominance; its place as the assumed and normal usage and meaning can hide things within these assumptions and this normality. Considering a dominant stability in terms of possible alternatives has the potential to expose some of those otherwise hidden qualities.[6]

I suggest that the idea in general of positive postphenomenological research into the multistability of technologies, and the ideas in particular of notions like "pivot points" and "dominant stabilities," all serve to highlight something important about technological mediation: its situatedness. Any understanding of multistability is necessarily context-relative. A choice of investigative pivot point is not made from a point of innocent objectivity (whether we're talking about a philosophical investigation of a technology's multistability, or a scientist's usage of this idea in empirical research). Such investigations are necessarily made from a particular subject position. This is why Whyte insists that postphenomenologists must put active work toward remaining as reflexive as possible, and that they should develop the interactional capabilities to engage others for whom this technology is relevant. The same is true for the notion of the dominant stability. This idea should immediately introduce the question: dominant for whom? When a postphenomenologist notices multiple stabilities for a given technology, what enables them to recognize the particular ones that they do? From what subject positions are a dominant stability the normal and assumed state of things?

This confrontation with situatedness is not a limitation of postphenomenological research; it is necessary part of the epistemology of philosophies of technology that maintain a commitment to a relational ontology. This is a political epistemology, one in which users and investigators are not innocent within their subject positions, and are always themselves co-constituted within their technological situation. There is a clear and obvious resonance here with the insights of feminist theories and others that have spent decades working to articulate these political dynamics in detail. This is consistent with the kinships noted in the previous section, as well as with the fellow traveler status that postphenomenology has long maintained with feminist theories of science. But there is work to do to follow out these implications for postphenomenological research.

5. A Funny Thing Happened on the Way to the Things Themselves

Ihde writes, "Husserl's call is for phenomenology to go to the things themselves. And technologies can, in a restricted sense, be things, or objects in an environment—only if they are sitting there, as it were, as objects not being used . . But such things do not present themselves that way in use" (2016, 130).[7] If we are to heed Husserl's call as philosophers of technology and go to the things themselves, then how are we to understand this project in light of our potential commitments to a relational ontology and technological multistability?

As revealed in many of the comments reviewed above, multistability is not an explanation of our technological situation. It is a finding. And it is a finding broadly consistent with the relational ontology to which postphenomenology and several other perspectives subscribe. It is a finding about the inherent relationality of our technological situation.

This can be considered in some respects a response to Husserl's call to go to the things themselves. We have. And what we have found has been surprising. The things themselves are such only in a multistable relationality with us, with other things, and with other people. The things themselves do not exist by themselves.

This introduces a number of questions and projects for postphenomenological research going forward. For example, what should we think of "invariants"? That is, one of the primary strengths of variational analysis, in both the Ihdean and Husserlian senses, is that it purports to get at something deeper about its object of investigation. The idea is that by approaching the object of study through multiple variations (or, put differently, by identifying multiple stabilities), we can reveal what is somehow essential or important to that object, and what it is instead merely contingent upon perspective. With regard to the relational ontology discussed here, we can wonder what we should think about the status of those features of the object of study. The answer is that even these "invariants" (those features of the object of study that are found to hold across all of the stabilities that we analyze) should themselves be understood as relative to the context of investigation (see esp. Ihde 1986, ch. 9, and Rosenberger 2017b). The discovery of invariants can provide elucidating information about the particular stabilities that are under study, teaching us something about their shared structures. However, if we are to remain consistent with a commitment to a relational ontology, then we should not further assume that such structural invariants tell us something independent of all context.

What should we think of the notion of essences? It is not my objective here to criticize or pass judgment on any specific account of essence on offer from other perspectives in phenomenology or the philosophy of technology, including Husserl's or Sartre's. However, it seems clear that any notion of "essence" within postphenomenological work should be used in a provisional manner, something that refers to a pattern that happens to hold across a particular context. One could remain roughly consistent with a relational ontology and, for example, do as Sartre does in the quote from "Existentialism Is a Humanism" and use essence to refer to the formula by which the designer makes the letter opener and the user immediately recognizes it as such. Whatever is being referred to there is something that is pervasive across some cultural context—the designer knows what those users will expect, and those users immediately do. "Essence," in this sense, ends up meaning something close to what has been referred to above as the dominant strategy. Crucially, if we are to be consistent with postphenomenology's commitment to a relational ontology, then the usage of essence in this sense does *not* refer to something fixed, or something deep and foundational. We see that despite anything we might call essence, the object of study is also open to additional stabilities. It does not appear that any conception of essence that holds it to be fixed and foundational can in any straightforward manner be made consistent with postphenomenology and other perspectives that maintain a commitment to a relational ontology.

All of this puts a spotlight on the epistemological situatedness of technological mediation. Within the inherent relationality of things, and within the context-relativity of our investigations into those things, we discover the political bearing of our own situatedness. Consistent with postphenomenology's integration of the commitments of American pragmatism, we find a blurring of the conventional distinctions between ontology and epistemology. The things themselves are not self-evident, but are made evident to an experiencer and within an experiential context, an experiential context made possible by the contingent specificities of technological mediation. In the search for the things themselves, and in the discovery of technological multistability, we find ourselves confronted with the politics of knowers and knowing. There is the opportunity here for postphenomenological research in particular, and philosophies of technology more generally, to connect up with feminist political epistemology, including standpoint theory (e.g., Harding 1986; Collins 1990; Hartsock 1998; Haraway 1997), epistemologies of ignorance (e.g., Sullivan and Tuana 2007), and issues of epistemic injustice (e.g., Fricker 2007; Dotson 2012).

We went back to the things themselves and discovered something unexpected: ourselves, there with those things.

Notes

[1] Due to the nature of the location of these characters, the stabbing has no effect.

2. Postphenomenology has grown mature enough to begin to draw multiple lines of critique. I like many of them. Relevant here are those suggesting that postphenomenology's focus on human-technology relations renders it unable to recognize the larger patterns—especially larger ethical and political patterns—regarding technology and society. There are also lines of critique that allege that for the same reason, postphenomenology is unable to make use of transcendental argumentation, which thus leaves it severely limited. Just a few of these include: Borgmann 2005; Scharff 2010; Smith 2015; Zwier et al. 2016; Coeckelbergh 2017, 180; Lemmens 2017. While these critiques are aimed mostly toward postphenomenology, they seem as though they should apply as well, at least to some degree, to related perspectives that rely on relational ontologies, including actor-network theory and feminist new materialism, although it may be a less popular move to criticize some of those potential targets.
   These criticisms are based on concrete things postphenomenologists have said and done. Verbeek has dismissed transcendental philosophies as backward-facing (2005). Ihde has spent much of his career criticizing any other philosophy of technology that makes overbroad or totalizing claims about technology, often suggesting that they fail to recognize technological multistability. Ihde's own general reluctance to engage in sustained ethical and political critique, combined with his penchant for criticizing others that do, has left postphenomenology open to the objection that its focus on human-technology relations leaves it without tools to engage in ethical and political work. The fact that these kinds of criticisms have also historically at times been leveled against the phenomenological tradition more generally gives us even more reason to expect them to continue to follow postphenomenology as well.
   My own general reaction to this growing body of critique is that it while it has merit, we must be careful not follow any further implication that technology has some kind of overarching and fixed essential nature. That is, I believe that we can find insight in these criticisms without abandoning a relational ontology and a commitment to technological multistability. For example, it is possible to develop a usage for transcendental argumentation that applies to limited spheres of technological phenomena and, crucially, refrains from assuming that results somehow obtain for all technology—whatever that even means.
   It should be noted as well that in addition to Verbeek's well-known postphenomenological work on technological ethics and the studies following in that vein (e.g., Verbeek 2011; Dorrestijn 2012; Kudina and Verbeek 2019), contemporary postphenomenological research has been demonstrating its potential for making unique and substantial contributions to political discourse and critique, as well as its potential to productively connect up with larger political perspectives (e.g., Warfield 2017; Wittkower 2017; Rosenberger 2017a; Rosenberger 2020; Verbeek 2020; Romele forthcoming).
3. For more on postphenomenology's relationship with pragmatism, see Ihde 2009; Ihde 2016; Rosenberger 2017b. Lenore Langsdorf's series of essays on this relationship and its implications are, in my view, essential reading on this topic (2015; 2016; 2020). For helpful considerations of this relationship offered by card-carrying pragmatists, see Mitcham 2006; Hickman 2008.
4. Or for example, as Ihde puts it: "This style of ontology carries with it a number of implications, including that there is a coconstitution of humans and their technologies. Technologies transform our experience of the world, and we in turn become transformed in the process" (2009, 44).
5. "Multistability," in its attempt to capture the idea that any technology can be used in different ways and can be developed differently along different trajectories, can be understood

to be one among a collection of related ideas in the fields of philosophy of technology and Science and Technology Studies. These include the notion of "interpretive flexibility" from the Social Construction of Technology perspective, the critical constructivism notion of technological "ambivalence," and even a conception of technology's potential status as a part of both a "program of action" and an "anti-program" in actor-network theory, among others (Pinch and Bijker 1984; Latour 1999; Feenberg 2017). Despite the fact that these ideas have sometimes been used interchangeably, I suggest that we should refrain from doing so since, in my view, these notions each capture subtly different and important aspects of the variability possible for technology. For example, while seemingly similar at first glance, the notions of multistability and interpretive flexibility—both decades-old ideas—help to articulate different phenomena. Where multistability is a phenomenological concept which refers to an ever-present potential for users to take up or interpret or develop a device differently, interpretive flexibility is a social concept referring to a status in which a device is interpreted differently by different groups.

6. Across a series of papers, I have developed a method called "variational cross-examination" for conducting the critical contrast of different stabilities that are identified for a given object of investigation. This process includes consideration of the material changes, bodily-conceptual approaches (what I've called "relational strategies"), and social and political enrollments distinctive to different stabilities (e.g., Rosenberger 2014; 2020; forthcoming; see also: Aagaard 2017). I've offered this method as a second step for postphenomenological investigations, one to follow Ihde's variational analysis.
7. As Lenore Langsdorf notes, "Telescopes, eyeglasses, and microscopes extend human vision; writing technologies supplement memory. However, Husserl's focus on 'the things themselves' did not take that expansion into account until his later work, most of which remained unpublished and thus not readily accessible" (2016, 116).

References

Aagaard, Jesper. 2017. "Introducing Postphenomenological Research: A Brief and Selective Sketch of Postphenomenological Research Methods." *International Journal of Qualitative Studies in Education* 30 (6): 519–533.

Aagaard, Jesper. 2018. "Magnetic and Multistable: Reinterpreting the Affordances of Educational Technology." *International Journal of Educational Technology in Higher Education* 15 (4): 1–10.

Aagaard, Jesper, Jan K. B. Friis, Jessica Sorenson, Oliver Tafdrup, and Cathrine Hasse, eds. 2018. *Postphenomenological Methodologies*. Lanham: Lexington Books.

Appleton, Charley. 2021. "Exploitable Multistability: The View from the Bike Lane." In *Postphenomenology and Architecture: Human Technology Relations in the Built Environment*, edited by Lars Botin and Inger Berling Hyams, 45–69. Lanham: Lexington Books.

Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*, 2nd ed. Durham: Duke University Press.

Bennett, Jane. 2010. *Vibrant Matter: A Political Ecology of Things*. Durham: Duke University Press.

Borgmann, Albert. 2005. "Review of *What Things Do*, by Peter-Paul Verbeek." *Notre Dame Philosophical Reviews*. Accessed January 8, 2005. https://ndpr.nd.edu/news/24832-what-things-do-philosophical-reflections-ontechnology-agency-and-design/.

Braidotti, Rosi. 2013. *The Posthuman*. Cambridge: Polity Press.

Braidotti, Rosi. 2019. "A Critical Framework for the Critical Posthumanities." *Theory, Culture & Society* 36 (6): 31–61.

Clark, Andy. 2010. *Supersizing the Mind*. New York: Oxford University Press.

Coeckelbergh, Mark. 2017. *Using Words and Things: Language and Philosophy of Technology*. London: Routledge.

Collins, Patricia Hill. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Boston: Unwin Hyman.

De Boer, Marjolein, and Jenny Slatman. 2018. "The Mediated Breast: Technology, Agency, and Breast Cancer." *Human Studies* 41: 275–292.

Dorrestijn, Steven. 2012. "Technical Mediation and Subjectivation: Tracing and Extending Foucault's Philosophy of Technology." *Philosophy & Technology* 25: 221–241.

Dotson, Kristie. 2012. "A Cautionary Tale: On Limiting Epistemic Oppression." *Frontiers* 33 (1): 24–47.

Feenberg, Andrew. 2017. *Technosystem: The Social Life of Reason*. Cambridge: Harvard University Press.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Fried, Samantha Jo. forthcoming. "Satellites, War, Climate Change, and the Environment: Are We at Risk for Environmental Deskilling?" *AI & Society*. DOI: doi.org/10.1007/s00146-020-01047-2.

Gallagher, Shaun. 2017. *Enactivist Interventions: Rethinking the Mind*. Oxford: Oxford University Press.

Haraway, Donna J. 1997. *Modest_Witness@Second_Millenium*. London: Routledge.

Haraway, Donna J. 2003. *The Companion Species Manifesto*. Chicago: Prickly Paradigm Press.

Harding, Sandra. 1986. *The Science Question in Feminism*. Ithaca: Cornell University Press.

Hartsock, Nancy C. M. 1998. *The Feminist Standpoint Revisited and Other Essays*. Boulder: Westview Press.

Hasse, Cathrine. 2020. *Posthumanist Learning: What Robots and Cyborgs Teach Us About Being Ultra-Social*. London: Routledge.

Hauser, Sabrina, Doenja Oogjes, Ron Wakkary, and Peter-Paul Verbeek. 2018. "An Annotated Portfolio on Doing Postphenomenology Through Research Products." *DIS '18*, June 9–13, 2018, Hong Kong. ACM, pp. 459–471.

Hayles, N. Katherine. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press.

Hickman, Larry. 2007. *Pragmatism as Post-Postmodernism*. New York: Fordham University Press.

Hickman, Larry. 2008. "Postphenomenology and Pragmatism: Closer Than You Might Think?" *Techné* 12 (2): 99–104.

Hoel, Aud Sissel, and Annamaria Carusi. 2015. "Thinking Technology with Merleau-Ponty." In *Postphenomenological Investigations,* edited by Robert Rosenberger and Peter-Paul Verbeek, 73–84. Lanham: Lexington Books.

Husserl, Edmund. 1900/1970. *Logical Investigations*, *vol. 1*, trans. J. N. Findlay. London: Routledge.

Ihde, Don. 1977. *Experimental Phenomenology: An Introduction*. New York: G. P. Putnam & Sons.

Ihde, Don. 1986. *Consequences of Phenomenology*. Albany: SUNY Press.

Ihde, Don. 1990. *Technology and the Lifeworld*. Bloomington: Indiana University Press.

Ihde, Don. 1993. *Postphenomenology: Essays in the Postmodern Context*. Evanston: Northwestern University Press.

Ihde, Don. 2009. *Postphenomenology and Technoscience: The Peking University Lectures*. Albany: SUNY Press.

Ihde, Don. 2016. *Husserl's Missing Technologies*. New York: Fordham University Press.

Keymolen, Esther. forthcoming. "In Search of Friction: A New Postphenomenological Lens to Analyze Human-Smartphone Interactions." *Techné*.

Kudina, Olya. 2019. "The Technological Mediation of Morality." Dissertation, University of Twente, the Netherlands. DOI: 10.3990/1.9789036547444.

Kudina, Olya, and Peter-Paul Verbeek. 2019. "Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy." *Science, Technology & Human Values* 44 (2): 291–314.

Langsdorf, Lenore. 2015. "Why Postphenomenology Needs a Metaphysics." In *Postphenomenological Investigations,* edited by R. Rosenberger and P.-P. Verbeek, 45–54. Lanham: Lexington Books.

Langsdorf, Lenore. 2016. "From Interrelational Ontology to Instrumental Ethics: Expanding Pragmatic Postphenomenology." *Techné* 20 (2): 112–128.

Langsdorff, Lenore. 2020. "Relational Ethics: The Primacy of Experience." In *Reimagining Philosophy and Technology, Reinventing Ihde*, edited by Glen Miller and A. Shew, 123–140. Dordrecht: Springer.

Lanng, Ditte Bendix, and Søren Risdal Borg. 2021. "Multistable Infrastructure: The Scripted and Unscripted Performance of a Functionalist Pathway." In *Postphenomenology and Architecture*, edited by Lars Botin and Inger Berling Hyams, 19–43. Lanham: Lexington Books.

Latour, Bruno. 1999. *Pandora's Hope*. Cambridge: Harvard University Press.

Lemmens, Pieter. 2017. "Thinking Through Media: Stieglerian Remarks on a Possible Postphenomenology of Media." In *Postphenomenology and Media*, edited by Yoni Van Den Eede, Stacey O. Irwin, and Galit Weller, 185–206. Lanham: Lexington Books.

Lewis, Richard S. 2021. *Technology, Media Literacy, and the Human Subject: A Posthumanist Approach*. Cambridge: Open Book Publishers.

Malafouris, Lambros. 2013. *How Things Shape the Mind: A Theory of Material Engagement*. Cambridge: MIT Press.

Mitcham, Carl. 2006. "From Postphenomenology to Pragmatism: Using Technology as an Instrument." In *Postphenomenology: A Critical Companion to Ihde*, edited by E. Selinger, 21–33. Albany: SUNY Press.

Moerenhout, Tania, Gary S. Fischer, and Ignaas Devisch. 2020. "The Elephant in the Room: A Postphenomenological View on the Electronic Heath Record and Its Impact on the Clinical Encounter." *Medicine, Healthcare & Philosophy* 23: 227–236.

Mozaffaripour, Roohollah. 2017. "Post-phenomenology as an Approach to Study Education Technology." *Journal of Foundations of Education* 6 (2): 63–81.

Pinch, Trevor J., and Wiebe E. Bijker. 1984. "The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." *Social Studies of Science* 14 (3): 399–441.

Romele, Alberto. forthcoming. "Technological Capital: Bourdieu, Postphenomenology, and the Philosophy of Technology Beyond the Empirical Turn." *Philosophy & Technology*. DOI: doi.org/10.1007/s13347-020-00398-4.

Rosenberger, Robert. 2014. "Multistability and the Agency of Mundane Artifacts: From Speed Bumps to Subway Benches." *Human Studies* 37: 369–392.

Rosenberger, Robert. 2017a. *Callous Objects: Designs Against the Homeless*. Minneapolis: Minnesota University Press.

Rosenberger, Robert. 2017b. "Notes on a Nonfoundational Phenomenology of Technology." *Foundations of Science* 22: 471–494.

Rosenberger, Robert. 2020. "'But That's Not Phenomenology!': A Phenomenology of Discriminatory Technologies." *Techné: Research in Philosophy and Technology* 24 (1/2): 83–113.

Rosenberger, Robert. 2021. "A Primer on Postphenomenology and Imaging." In *Postphenomenology and Imaging: How to Read Technology*, edited by Samantha J. Fried and Robert Rosenberger, forthcoming. Lanham: Lexington Books.

Rosenberger, Robert. forthcoming. "On Variational Cross-Examination: A Method for Postphenomenological Multistability." *AI & Society*. DOI: doi.org/10.1007/s00146-020-01050-7.

Rosenberger, Robert, and Peter-Paul Verbeek, eds. 2015. *Postphenomenological Investigations: Essays on Human-Technology Relations*. Lanham: Lexington Books.

Sartre, Jean-Paul. 1946a/1975. "Existentialism Is a Humanism." In *Existentialism: From Dostoevsky to Sartre*, edited by Walter Kaufmann, 345–369. London: Plume.

Sartre, Jean-Paul. 1946b/1989. *No Exit and Three Other Plays*. New York: Vintage International.

Scharff, Robert C. 2010. "Technoscience Studies after Heidegger? Not Yet." *Philosophy Today* 54: 106–114.

Shaw, Sara E., Gemma Hughes, Sue Hinder, Stephany Carolan, and Trisha Greenhalgh. 2020. "Care Organizing Technologies and the Post-phenomenology of Care: An Ethnographic Case Study." *Social Science & Medicine* 255 (112984): 1–9.

Sicart, Miguel. 2020. "Playing Software: The Role of the Ludic in the Software Society." *Information, Communication & Society* 23 (14): 2081–2095.

Smith, Dominic. 2015. "Rewriting the Constitution: A Critique of 'Postphenomenology.'" *Philosophy & Technology* 28: 533–551.

Sullivan, Shannon, and Nancy Tuana, eds. 2007. *Race and Epistemologies of Ignorance*. Albany: SUNY Press.

Vallor, Shannon. 2015. "Beyond Ordinary Givenness? Postphenomenology, Digital Imaging, and Evidentiary Responsibility." In *Technoscience and Postphenomenology: The Manhattan Papers*, edited by Jan K. B. O. Friis and Robert P. Crease, 19–37. Lanham: Lexington Books.

Van Den Eede, Yoni, Stacey O. Irwin, and Galit Wellner, eds. 2017. *Postphenomenology and Media: Essays on Human-Media-World Relations*. Lanham: Lexington Books.

Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*, trans. Robert P. Crease. University Park: Penn State University Press.

Verbeek, Peter-Paul. 2011. *Moralizing Technology*. Chicago: Chicago University Press.

Verbeek, Peter-Paul. 2020. "Politicizing Postphenomenology." In *Reimagining Philosophy and Technology, Reinventing Ihde*, edited by Glen Miller and Ashley Shew, 141–155. Dordrecht: Springer.

Wakkery, Ron. forthcoming. *Things We Could Design: In More than Human-Centered Worlds*. MIT Press.

Warfield, Katie. 2017. "MirrorCameraRoom: The Gendered Multi-(In)Stabilities of the Selfie." *Feminist Media Studies* 17 (1): 77–92.

Wellner, Galit. 2020. "The Multiplicity of Multistabilities: Turning Multistability into a Multistable Concept." In *Reimagining Philosophy and Technology, Reinventing Ihde,* edited by Glen Miller and Ashley Shew, 105–122. Dordrecht: Springer.

Whyte, Kyle Powys. 2015. "What Is Multistability? A Theory of the Keystone Concept of Postphenomenological Research." In *Technoscience and Postphenomenology*, edited by Jan Kyrre Berg Olsen Friis and Robert P. Crease, 69–81. Lanham: Lexington Books.

Wiltse, Heather. 2020. "Revealing Relations of Fluid Assemblages." In *Relating to Things: Design, Technology, and the Artificial*, edited by Heather Wiltse, 239–253. London: Bloomsbury.

Wittkower, Dylan E. 2017. "Discrimination." In *Spaces for the Future: A Companion to the Philosophy of Technology*, edited by Joe Pitt and Ashley Shew, 14–28. New York: Routledge.

Zwier, Jochem, Vincent Blok, and Pieter Lemmens. 2016. "Phenomenology and the Empirical Turn: A Phenomenological Analysis of Postphenomenology." *Philosophy & Technology* 29: 313–333.

PART V

TECHNOLOGY, AESTHETICS, AND DESIGN

CHAPTER 20

UNDERSTANDING ENGINEERING DESIGN AND ITS SOCIAL, POLITICAL, AND MORAL DIMENSIONS

PHILIP BREY

1. Introduction

The philosophy of (engineering) design has emerged in recent decades as a focal area in the philosophy of technology (Vermaas and Vial 2018; Parsons 2015). On the one hand, it has attracted the attention of philosophers after the empirical turn in the philosophy of technology (Kroes and Meijers 2001; Brey 2010a), who hold that a philosophical understanding of engineering design is vital for a philosophical understanding of technology and its consequences for society. On the other hand, many designing engineers are interested in reading about, and contributing to, philosophical discussions of their core practice.

The philosophy of design is concerned with the nature of design, its central concepts, assumptions, theories, and methods; its relation to other human practices; its role in society; and its social, moral, cultural, and political dimensions. In analytic philosophical traditions, there is a focus on understanding and analyzing the concepts, methods, assumptions, practices, and products of engineering design (Vermaas et al. 2008; Kroes 2012; Meijers 2009; Chakrabarti and Blessing 2014). In continental approaches, the focus is on philosophical-anthropological and social-philosophical analyses of the role and significance of design for humans and society, as well as its aesthetic, cultural, and transcendental dimensions, and there is often a focus on design in general, rather than engineering design alone, with special attention to industrial design, interaction design, architecture, and graphic design (Willis 2018; Bardzell et al. 2018). In both traditions, there have been efforts to address the role of values and politics in design and to investigate ways of introducing ethical, social, and political considerations into design (van den Hoven et al. 2015a; Verbeek 2011).

My main interest is in the moral, social, and political implications of design. How do designs and design processes include implicit moral, social, and political choices that affect society? How can we explicate these choices and amend design processes as a result to make them good designs that are good in an ethical sense and good for society? This will be the main focus of this chapter. However, before we get to a detailed analysis of the relation between engineering design and society, I believe we should first have a proper understanding of engineering design itself, including its nature, structure, and function, its relation to other human practices, and the different types of engineering design that exist. In the next section of this chapter, therefore, I will give an account of engineering design. This account draws from both philosophical studies of engineering design and accounts from within engineering itself. The core of this section is an account of the structure of engineering design processes that will subsequently be used in my account of the moral, social, and political implications of design.

The section that follows focuses on the moral, social, and political implications of design. I will investigate what a good design is from the perspective of ethics and society, how new designs can affect society in positive and negative ways, and how design processes can be supportive of values and ideals of a good society. I will do so in reference to studies of embedded values in design, approaches for the incorporation of values and ethics into design, and theories of the social and political dimensions of design.

2. What Is Engineering Design?

This section will concern the question of what engineering design is and how it is structured. I will begin by answering the question of what type of practice engineering design is, and how it is distinct from other types of design and other human practices. I will then proceed to situate engineering design within the practice of technology development and engineering at large, and will consider its role within, and relation to, innovation. I will conclude by examining the structure of engineering design processes, and how these feed into production and marketing.

2.1 Engineering Design and Other Forms of Design

Designing is the creation of a plan for the construction or realization of an object, system, process, or feature. This plan can be of different types: it can be a description of the entity that is to be realized, a series of instructions, a drawing, a graphical model, a series of mathematical equations, or yet something else. Designing is a core activity in a number of professional fields: those fields that are concerned with the planning and production of new things, systems, and processes. Design, in these fields, encompasses the stage during which plans are made for the production of these new things. These fields include the following:

- Engineering, in which one of the central activities is the design of new technological artifacts, systems and processes
- Craft and applied arts (pottery, ceramics, graphics, metal works, textile arts, interior design, etc.)
- Fine arts (painting, sculpture, photography, music, etc.)
- Architecture

Sometimes, “design” is also used in relation to certain branches of the applied social sciences, and then it refers to the planning of new social structures, practices, or events. However, the term “design” is used less frequently in these fields, and instead words like “planning” and “modeling” are more often used. Nevertheless, there are professional activities in the applied social sciences, in which “design” is a central term, like organizational design (the improvement of organization structures and processes to better fit organizational objectives), social design (the design of social structures and processes in order to help solve social problems and promote human welfare),[1] and communication design (the planning and shaping of messages in content, form, and delivery channels).

The word “design” is also used for planning activities by professionals who are not necessarily applied social scientists but who nevertheless make plans for new activities, events, social structures, forms, or organizations, as when a teacher is said to design a new curriculum, or when an administrator designs a new form. And finally, the word “design” is also used in reference to everyday activities of planning, as when it is claimed that a person has designed a plan for making new friends, a cozy reading corner in their home, or a system for distributing and tracking household chores.

Design is therefore an activity that is much more encompassing than engineering design alone. It is a core human activity even in societies that are not technologically advanced. We are *homo faber*, beings that make things, and part of our success as a species is that we use our intellect to develop plans for new tools, artifacts, practices, social arrangements, and other new things that we consider to be useful or meaningful. The activity of making plans or blueprints for such new things is called *designing*, and the plans themselves are *designs*. Designs are usually inscribed in an external medium that people can use as a model or set of instructions for realizing the design. This external medium can be a document, a picture, a physical model, or some other type or representation or series of instructions. Designs can also be internalized, in the mind, as when someone has devised a plan for a new artifact but has not yet put it on paper. Designs can sometimes also be read from things produced that are based on them. When someone has knitted a sweater with an interesting new pattern, for example, people need not see a separate plan for the sweater to understand the new design, as it is in plain view for them.

Engineering design can be distinguished from other types of design by considering the special nature of the activities that it involves. The American Accreditation Board for Engineering and Technology (ABET) defines engineering design as "the process of devising a system, component, or process to meet desired needs. It is a decision-making process (often iterative), in which the basic science and mathematics and engineering sciences are applied to convert resources optimally to meet a stated objective" (ABET 2018). ABET moreover defines engineering as "the profession in which a knowledge of the mathematical and natural sciences gained by study, experience, and practice is applied with judgment to develop ways to utilize economically the materials and forces of nature for the benefit of mankind" (ABET 2018).

These definitions also underline what scholars in technology studies have claimed about engineering design: that it is a form of design that relies on specialist training in engineering science, which includes extensive knowledge of the mathematical and natural sciences, and the methods of applying such knowledge. The intense application of mathematics and natural sciences is certainly something that sets engineering design apart from other forms of design. However, as many scholars have argued, engineering is not merely the application of science and mathematics; it also involves the creation and application of unique engineering knowledge, which is a highly formalized, evidence-based, and systematic type of knowledge (Vincenti 1990). Based on these studies, a more adequate definition of engineering design than the ABET definition would state that design involves the application of science, mathematics, and engineering knowledge. So let us reformulate the definition of engineering design: Engineering design is the development, through the application of science, mathematics, and engineering knowledge, of plans for products (devices, systems, methods, procedures) that can serve practical ends.

Engineering design is not only distinct in its practices, but also in its resulting plans. As Clive Dym (1994) has argued, engineering design uses special "languages of design," that is, particular symbol systems, notations, systems of icons, and graphical conventions for drawing up and communicating design plans, that are altogether different from those used in other fields. These languages are used to represent objects and processes. As Dym claims, designers use a physical representation language based on mathematics, science, and engineering knowledge to produce mathematical and analytical models to express some aspect of an artifact's function or behavior. Designers also use graphical representations of various kinds, often involving exact measurements and representational conventions from the engineering sciences, and often interpreted within CADD systems. In addition, designers use verbal or textual statements to document and communicate designs and describe objects, constraints, and limitations with concepts and terms unique to the engineering sciences and symbolic representations derived from symbolic computing and AI-based programming, such as if–then rules, frames, and computationally defined objects.

Engineering design is also distinct in the products and processes that are realized on the basis of designs, which have unique characteristics not found in the products of other types of design. To demonstrate how this is so, I will consider the four main branches of engineering and the designs that they typically yield. The four main branches of engineering are chemical engineering, civil engineering, electrical engineering, and mechanical engineering. They are involved in the design and manufacture of artifacts that typically cannot be produced outside of engineering.

Chemical engineering is involved with the production, transformation, and utilization of chemicals, materials, and energy through the application of principles of chemistry, physics, and mathematics. Chemical engineering enables the production of artifacts such as medicine, petrochemicals, and plastics, and the development of processes such as oil refinery and mineral processing, all of which could not exist if it were not for chemical engineering design.

Mechanical engineering is concerned with the design, analysis, manufacture, and maintenance of mechanical systems. It applies physics, mathematics, materials science, and engineering knowledge to do so. It may be observed that mechanical systems like steam engines, windmills, and water wheels were already developed and used thousands of years ago, before the development of the engineering sciences as we currently know them. Although such relatively simple mechanical systems can be developed outside of mechanical engineering, it is only the application of sophisticated science, mathematics, and engineering knowledge in mechanical engineering that has enabled the production of more advanced mechanical systems such as engines, automobiles, industrial machines and robots, and the optimization of simpler systems like windmills and steam engines.

Electrical engineering is concerned with electrical, electronic, and electromagnetic systems—such as televisions, telephones, radar systems. and the electric grid—that clearly could not be manufactured if it were not for electrical engineering and its extensive reliance on mathematics and natural science.

Civil engineering is concerned with the design, construction, and maintenance of the built environment, including structures such as roads, bridges, canals, dams, sewerage systems, and railways. Many of these structures were already being designed and built by artisans thousands of years ago. The emergence of a scientific approach in civil engineering, however, has led to dramatic advances in the kinds of structures that can be built and the functionality that they have.

Computer science and computer engineering are more recent fields that do not neatly fit within this engineering taxonomy. Computer science is normally considered to be a branch of science rather than engineering. It is the study of computing devices and the way in which they process, store, and communicate data and instructions. This is often done toward a practical end, however, which is to improve the processing of data and instructions in computing devices. Because of this practical aim, computer science has a resemblance to engineering, even if it is considered a science. Computer engineering is different from computer science. It is the combination of computer science and electrical engineering. It is generally considered to be a branch of electrical engineering.

Computer engineers design computer hardware and software, as well as systems that integrate both.

2.2 Situating Engineering Design in Engineering and Innovation

Engineering design is a central practice in engineering. Yet, it is not the only practice. Engineers are also involved in research activities prior to design and in activities that take place after design, notably the manufacturing, operation, and maintenance of technological artifacts and systems. Research in the engineering sciences is distinct from research in the natural sciences in that it is, to a greater or lesser extent, application-oriented (Boon 2011). It follows scientific methods, including scientific methods of experimentation, observation, hypothesis testing, and establishment of law-like relationships, but its aim is not to uncover perennial truths about the universe, but rather to create useful knowledge that may have a future application in engineering design. Examples of such research include the investigation of properties of different types of alloys in materials science and the study of the impact of liquid droplets on superheated surfaces. Sometimes the term *engineering scientist* is used to designate engineers who engage in this type of applied research.

Engineers also have roles in production and manufacturing. Notably, such roles are taken up by production engineers. Production engineers are involved in the design of equipment, tools, and machinery used in manufacturing processes, as well as in the implementation, monitoring, and optimization of manufacturing and production processes. They work together with many other professionals in the production and manufacturing process who often do not have engineering degrees, such as assemblers, machinists, welders, production managers, and quality control inspectors. Engineers can have roles in maintenance, as well. Maintenance engineers are involved in the checking, repairing, and servicing of machinery, equipment, systems, and infrastructures. As these cases show, there are many engineering professions and jobs in which engineering design is not central. However, engineering design, being the activity aimed at inventing, defining, and planning the technological artifacts and processes, is clearly a salient and central component of engineering.

As engineering design is central in engineering, it is also central in technological innovation. *Technological innovation* is the invention of new concepts, techniques, and designs in engineering that are then realized into products and subsequently marketed and included in social and economic practice. Technological innovation is more than mere *invention*, which is merely the development of new ideas, concepts and designs (Malerba and Orsenigo 1997). It goes beyond invention by requiring implementation: it also involves subsequent product realization, marketing, and diffusion into society. Although technological innovation often depends on innovative designs, it can also be the result of the invention of new concepts and techniques at research and pre-design stages, and can also involve innovative production and marketing processes.

Technological innovation is only one type of innovation. 'Innovation' can be defined as activities by an organization or unit to produce innovations, and an innovation is "a new or significantly improved product or process (or combination thereof) that differs significantly from the unit's previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process)" (OECD/Eurostat 2018, 20). An innovation can be a technological product or process but also a regular good or service; a new marketing method or commercial practice; a new policy; or a new organizational method, form, or practice. Innovation can be undertaken by commercial firms but also by governments, NGOs, and other organizations and groups. Innovation undertaken to better meet social needs is called *social innovation*. As seen in the definition of innovation provided, a distinction is often made between 'product innovation' (the introduction of goods or services that have new or improved characteristics or uses) and 'process innovation' (the implementation of new or improved production or delivery methods). This distinction also applies to technological innovation.

It should be observed that not all technological design is necessarily innovative. Much technological design is routine design. 'Routine design' is defined by Gero (1990, 32) as "design that proceeds within a well-defined state space of potential designs. That is, all the variables and their applicable ranges, as well as the knowledge to compute their values, are all directly instantiable from existing design prototypes." Routine design does not involve much innovation and creativity. At the other extreme, one finds 'innovative' and 'creative design,' which involve substantially new design plans or solution principles, and in between are various forms of 'redesign,' including variant and adaptive design, in which an existing design is improved upon by finding ways to satisfy new requirements or improve performance (Pahl and Beitz 1996).

2.3 Structure of the Engineering Design Process

In theoretical and methodological studies of design, in engineering design textbooks, and to a lesser extent in the philosophy of design, considerable attention is paid to the structure of the design process. In accounts of this structure, various steps or phases of design practice are distinguished and related to each other, often with elaborate diagrams to illustrate the different steps. Most authors distinguish four to eight stages in design, which often can be iteratively applied, starting from formulation of the problem or need and formulation of design requirement, to conceptual design, in which basic ideas are formed for the solution to the problem, including the broad outlines of function and form, to detailed design, in which detailed plans, specifications and cost estimates are made, and in which final instructions are made for production (Johanneson and Perjons 2014; Jack 2013; Chakrabarti and Blessing 2014).

I focus here on the account of design processes provided in a prominent study of engineering design by Pahl, Beitz, Feldhusen, and Grote (2007). Pahl et al. describe the design process as having five phases:

1. *Product planning* is the development of an idea for a new product that results in a task description for an engineering department for development of the new product. Product planning is often not done by designers themselves but by clients and product planning departments or marketing departments of companies. It is often based on a real or perceived need expressed by a client or thought to be located in the market.
2. *Task clarification* is the process of clarifying the kind of product that is needed, identifying and formulating requirements and constraints, and creating a list of requirements, or design specification. Product planning and task clarification are often integrated processes in which there is a movement back and forth between planning and clarification.
3. *Conceptual design* is the process of finding solutions to any problems posed by the design specification at a conceptual level. Conceptual design involves identifying essential problems through abstraction, establishing function structures in which overall functions are divided into subfunctions, searching for appropriate working principles to drive the subfunctions, and combining them into working structures. The result is called a design concept or principle solution.
4. *Embodiment design* is a phase in which a design concept is developed into a definitive layout of the proposed technical product or system. This involves developing a layout design that defines the general arrangement and spatial features of the product, a preliminary form design that stipulates component shapes and materials and production processes, as well as providing solutions for any auxiliary functions not covered in the conceptual design stage. It strongly involves technical and economic considerations, and must result in a design that can be checked for its function, durability, production and assembly, operation, and cost. Embodiment design often involves several repeat design processes before a definitive design emerges.
5. *Detail design* is a phase that completes the embodiment design process with final instructions before production. These final instructions concern shapes, forms, dimensions, and surface properties of components; a definitive selection of materials; a final specification of production methods, operating procedures, and costs; and the development of production documents that include component and assembly drawings and parts lists. This is still done by design departments rather than production departments. Detail design may also involve the development of assembly instructions, transportation documentation, and quality control measures for the production department and operating, maintenance, as well as repair manuals for users.

Pahl et al. emphasize that engineering design is an iterative process: at any phase in the design processes, designers may retreat to an earlier phase, and it is also possible that different engineering teams work on different phases simultaneously.

After the detail design phase, the production department takes over from the engineering department and manufactures the product. In practice, detail design and production often overlap and thus require close collaboration between design and production departments. After production, there is transfer to the client and/or installation (for unique products) or marketing (for mass-produced products). For mass-produced products, user and marketing analytics, which is increasingly based on big data analytics, will often be collected after distribution and consumption, which could then lead to changes in the design for new batches of the product (Eppinger and Geracie 2013; Xu et al. 2016).

A potential weakness of the Pahl et al. account is that it makes little reference to prototyping and testing, processes that are often used in engineering design. 'Prototyping' is the production of inexpensive, scaled-down versions of a product or specific features of it, so that problem solutions generated at an earlier stage can be investigated. Pahl et al. do cover its role in design, but only briefly. They claim that prototyping can occur at any stage in the design process and that it frequently is used at the conceptual stage to test fundamental design concepts, but also at later stages in the design process (Pahl et al. 2007, 133). *Testing* is the assessment of the performance, safety, quality, or compliance with standards of a designed product or system, subsystem, or component. Testing can be done through prototyping, but is often done with a fully realized product, subsystem, or component. Consumer testing is a special form of testing, in which the product is tested with prospective consumers to see if it meets their expectations. Testing often takes place during production, after which results can feed back into design if the test results give indication that a redesign is needed. It also takes place during the design process, however, where it can occur during almost any stage, but especially during the later stages. It seems to be a weakness of the Pahl et al. account that it makes very little reference to testing.

3. Good Design and the Ethics of Design

In this section, I investigate to what extent and how moral, social, and political choices are embedded in design and how they can be designed for. I start by investigating what it means to say that a design is good, and I examine the relation between engineering design, on the one hand, and values, benefits, and the good of society, on the other hand. Then, in section 3.2, I investigate how consequences for society can be embedded in design, and in section 3.3, I conduct a parallel investigation of the embedding of values in design. Finally, in section 3.4, I consider approaches to designing for values and benefits to society.

3.1 What Is a Good Design?

A good design is a design that results in a good technological product. So what, then, is a good technological product? One answer is that it is a product that fulfills its function well. On this conception, a good microwave oven is one that is good at microwaving food, and a good radar system is one that is good at detecting moving and stationary objects. Let us call this type of goodness 'functional goodness.'[2]

A second answer is that a good technological product is one that is good at meeting the design requirements that have been specified for it. For example, the design requirements for a wrist watch may include requirements such as ability to tell the time (its proper function), being made out of metal parts, being of certain dimensions so as to be wearable, being made of nontoxic materials, being original in its design, being cost-effective to make, being easy to read, not having sharp edges, and being able to be mass-produced. Let us call this type of goodness 'requirements goodness.' Note that requirements goodness normally includes functional goodness: among the requirements for a new technological design are usually requirements that one or more functions are performed well by the product in question.

A technological product may be good in the requirements or functional sense, but still be bad in other ways. For example, a product may be bad for one's health or bad for the environment despite having functional and requirements goodness. This can happen when its original requirements do not include those of it not being harmful to health or to the environment. This type of goodness, when something is not good or bad *at* something (such as performing a function or meeting requirements), but good or bad *for* something, is called 'prudential goodness' (Fletcher 2012). It is a relation between an entity *E* and an entity *F* for which or whom *E* is good.[3] To say that *E* is good for *F* is to say that *E* contributes to the existence, flourishing, welfare or excellence of *F*. *F* can be anything that is of positive (intrinsic or instrumental) value. In particular, it can denote persons, positive or desired conditions, qualities or capabilities of persons (e.g., health, [low] blood pressure, endurance), groups (e.g., children, disabled individuals), practices and institutions (e.g., the economy, family life), social conditions and values (e.g., social cohesion, civility, privacy), as well as the environment and society at large.

The types of prudential goodness that have traditionally been considered to be the most important are goodness for persons and goodness for society. Other types of prudential goodness are arguably subordinate or contributory to these two more fundamental types. For example, goodness for health is contributory to, and subordinate to, goodness for persons, since that things go well for us in general is more important to us than things going well with our health, because the latter state does not prohibit other things for us going badly. Likewise, goodness for the economy is contributory to, and subordinate to, goodness for society.

In Brey (2018) I argue that the goodness of society is more important than goodness for persons, since the well-being of persons should be seen as a component of any conception of a good society. I moreover argue that next to well-being, justice is an intrinsically valuable good in society, and that other dimensions of a good society, like democracy, freedom, sustainability, and community, are best analyzed as instrumentally valuable to well-being and justice. There are, however, different conceptions of a good society, in which for example sustainability or ecological integrity is seen as intrinsically valuable, or in which democracy, autonomy, or individual rights are seen as intrinsically valuable. On many theories of goodness, however, goodness for society, however it is conceived, is the most important or highest form of goodness, and therefore the highest form of goodness for a technological product is its goodness for society. This means that a prudentially good design, in the most general sense, is one that results in products that tend to be good for society.

Prudential goodness (for society, human beings, or something else) is not the same as moral goodness, and in philosophy, the two have usually been distinguished. Moral goodness relates to right and wrong. Someone else's money can be prudentially good for me, but it can be morally wrong for me to accept it if it is not freely given. Prudential and moral goodness can, however, be related in the following way. Moral values are among those things that can be benefited or harmed, as when one says that actions harm privacy or support justice. So entities can be prudentially good or bad for moral values. A technological product can therefore be said to be morally good if it is prudentially good for moral values. A technological product is morally good for moral value *V* if it tends to support *V* rather than violate it. For example, Internet software that tends to divulge one's personal information to third parties is morally bad with respect to privacy, and software that tends to support the protection of personal information is morally good with respect to privacy. When a technological product tends to support all key moral values, we can say that it is morally good in a general sense.

Moral goodness is, in my view, contributory to the goodness of society. That is, a society in which people behave morally, institutions are arranged in accordance with moral principles, and technological products tend to be supportive of moral values is a better society than one in which this is not the case. It should also be clear, however, that moral goodness is not constitutive of the overall goodness of society. That is, there is more to being a good society than it being a moral society. A society can be moral, but still fall short because it has a poor economy, poor institutional arrangements, poor management of hazards and risks, and other shortcomings that keep it from being a good society. In the view I am proposing, one of the ways in which technological products can contribute, or fail to contribute, to the goodness of society is through their upholding, or violation, of moral values, and when a technological product upholds a moral value one could say that it is prudentially good for that value, and thereby, at least with respect to its support of morality, that it is prudentially good for society.

In conclusion, we have learned that a design (of a technological product) can be called good in at least four senses: it can be functionally good, have requirements goodness, be prudentially good, and be morally good. The last type can, however, be subsumed as a special kind of prudential goodness. The most important form of prudential goodness that can be considered in design is goodness for society, as it arguably encompasses other forms of prudential goodness, including goodness for moral values (moral goodness). I have argued that goodness for society appears to be a more important form of goodness for technological products than functional or requirements goodness. I now turn to the question of whether and how both goodness for society and moral goodness can be considered in design. I do so by examining how designs may affect society and how designs may affect the realization of moral and non-moral values, after which I will consider how these influences may be accounted for in design.

3.2 Technological Products with Built-in Consequences

The question is then whether we can come up with a viable theory of technological design according to which designs can yield products that are in a systematic and predictable way contributory to the goodness of society and its constituent parts. A possible argument against the existence of a viable theory of this sort is that that it is the use of an artifact that determines its effects, not the design. I have called this the *neutrality thesis* (Brey 2010b): the thesis there are no consequences that are inherent to technological artifacts, but that artifacts can always be used in a variety of different ways, and that each of these uses comes with its own consequences.[4] The neutrality thesis can be made plausible with examples of simple tools like hammers and razors. A hammer can be used to hammer nails, but also to break objects, to kill someone, to flatten dough, to serve as a paper weight or to conduct electricity. Different uses of a hammer have radically different effects on the world, and there do not seem to be single effects constant in all of them. If the neutrality thesis is true, it would seem to follow that attempts to improve society should perhaps not pay much attention to technological artifacts themselves, because they in themselves do not "do" anything. Rather, they should focus on the usage of these artifacts.

As many have argued, however, the neutrality thesis is false (Rose 2012; Verbeek 2005; Brey 2010b). Cases to buttress the neutrality thesis usually make reference to versatile tools like hammers, which have many very different uses. Most technological products, however, have only a limited range of (sensible) uses, and there are recurrent consequences across many or all of these uses. An ordinary gas-engine automobile, for example, can evidently be used in many different ways: for commuter traffic, for leisure driving, to taxi passengers or cargo, for hit jobs, for auto racing, as a temporary shelter for the rain, or as a barricade. Whereas there is no single consequence that results from all of these uses, there are several consequences that result from a large number of these uses: in all but the last two uses, gasoline is used up, greenhouse gases and other pollutants are being released, noise is being generated, and at least one person (the driver) is being moved around at high speeds.

These uses also have something in common: they are all central uses of automobiles, meaning that they are accepted uses that are frequent in society and that account for the continued production and usage of automobiles. The last two mentioned uses are peripheral in that they are less dominant uses that depend for their continued existence on these central uses, because their central uses account for the continued production and consumption of automobiles. Central uses of automobiles make use of their capacity for driving, and when it is used in this capacity, certain consequences such as the ones mentioned are very likely to occur. What this example suggests is that technological products are not neutral but may be claimed to have cross-cutting, "embedded" or "built-in" consequences or effects. What this means is that particular consequences manifest themselves in all of the *central* uses of the technological product (Brey 2010b). A central use is a use that is prevalent in society, and tends to make use of advanced functional features of the product, that are the result of a complex technological design.

It should be acknowledged that even if a technological product is used according to one of its central uses, there are often ways to avoid particular consequences. For example, a gas-fueled automobile need not emit greenhouse gases into the atmosphere if a greenbox device is attached to it, which captures carbon dioxide and nitrous oxide and converts it into bio-oil. The notion of a built-in consequence does not refer to consequences that are necessary and unavoidable, but rather to strong tendencies. So no strong technological determinism is implied, but only a weak, contextual determinism, which holds that technological products can be associated with recurrent effects that have a tendency to manifest themselves across their central uses, barring exceptional circumstances (Brey 2005). To deny such recurrent effects is to fall back into the neutrality thesis and therefore to miss the opportunity to address these recurrences in the design process. It is simply wrong to say that the emission of greenhouse gases by automobiles is a result of their use and not their design, when there are designs that are associated with such emissions (as in gasoline-fueled cars) but also designs that are not (as in electric cars). In Brey (2006), I present a taxonomy of different kinds of recurrent consequences of technological products, including social, cultural, material, behavioral and other types of consequences.

I have argued previously (Brey 2005) that recurrent effects associated with technological artifacts can be understood as resulting from affordances and constraints associated with an artifact. Affordances are new actions, events or configurations of the environment opened up by artifacts. Constraints are limitations to configurations of the environment imposed by artifacts. Embedded consequences of technological products can moreover often be evaluated as positive or negative. If they are evaluated as positive, they may be called embedded or built-in *benefits*. For example, Bruno Latour's (1990) hotel keys with a weight attached have as a benefit that they tend to be deposited at the front desk. If embedded consequences are negative, they are embedded *harms*. For example, the emission of greenhouse gases is an embedded harm of gas-engine cars.

3.3 Technological Products with Built-in Values

We have seen that technological products can be associated with "built-in" consequences, and that these consequences can be beneficial or harmful in relation to persons and other valuable entities. I will now claim that just as technological products can be beneficial or harmful to persons, the economy, or the environment, they can also be beneficial or harmful to *values*. That is, they can be beneficial or harmful to the realization of values in the real world, meaning the extent to which events and states-of-affairs are shaped or brought into effect in accordance with particular values. Freedom, justice, or privacy are abstract qualities, of which there can be more or less in the world. The amount of freedom in the world, for example, depends on the extent to which individuals have freedom of movement, thought, expression, and association. If many individuals do not have this, there is less freedom in the world, and if many have it, there is more. For a technological product to be beneficial to freedom, therefore, it must have a systematic tendency, across different uses, to bring about more freedom in the world.

The claim I want to make, then, is that technological artifacts can have systematic tendencies to promote or benefit values such as privacy and sustainability, as well as tendencies to harm or detract from them. In short, one can associate technological products with values embedded in them. This approach to technology is called the 'embedded values approach' (Nissenbaum 1998). Observe that, following from the definition of prudential goodness in section 3.1, a technological product that promotes or upholds a value is prudentially good for (the realization of) that value, and one that harms a value is prudentially bad for it. So a product with an embedded value of privacy is (prudentially) good for privacy, and one with an embedded tendency to harm privacy is (prudentially) bad for privacy. The embedded values approach was originally formulated by Helen Nissenbaum (1998; Flanagan et al. 2005) and Batya Friedman (Friedman et al. 2006). I have also worked on an embedded values approach since the late 1990s (Brey 2000, 2010b).

An approach related to the embedded values approach, and chronologically preceding is, is the approach of embedded politics in technological products. Langdon Winner (1980) famously asked, "Do artifacts have politics?" and then proceeded to answer this question affirmatively. The politics of artifacts can concern their promotion of particular political arrangements and processes (e.g., hierarchical structures, privatization processes), but also political values and ideals (e.g., distributive justice, democracy, equality). If the latter are at issue, then the embedded politics approach coincides with the embedded values approach. Another related approach is the technomoral virtue ethics approach of Shannon Vallor (2016). Vallor claims that particular technologies tend to promote the development of certain virtues and vices in users: virtues such as honesty and empathy, and vices such as dishonesty and carelessness. This approach can also be understood as a special version of the embedded values approach.

3.4 Designing for Values, Benefits, and a Good Society

The idea that technology is not neutral and that values and consequences can, to some extent, be embedded in design, is at the heart of various approaches to design that have been developed in recent decades. I first briefly consider approaches to design that focus on the realization of certain types of benefits, or that focus for benefits for society at large, after which I discuss approaches that focus on the realization of values.

There are many approaches to design that focus on the realization of particular benefits for society. Environmental design is an approach to design that focuses on developing products and structures that are sustainable and beneficial for environment and health. User-centered design is design that tries to better accommodate for the needs, goals, and behavioral tendencies of users. Universal design is the design of product and environments all people, without the need for adaptation or specialized design. Behavioral design (Wendel 2013) and persuasive technology (Fogg 2002) are approaches that aim to change people's behavior, daily routines, and thinking, thereby providing benefits to users and society. Social design (Sachs et al. 2018) is design aimed at solving social problems, improving welfare, and bringing about social change. In these approaches, the social benefit that is being designed for can either be encoded in the proper function of products (e.g., a weight-loss app that has the function of influencing food intake, a waste-sorting system that has the function of enabling recycling, and hence contributes to sustainability) or be an embedded benefit distinct from the proper function (e.g., an electric car, whose function is transportation, but that also contributes to sustainable practices).[5]

Design approaches based on the concept of embedded values find their beginning in the seminal work of Batya Friedman and her associates (Friedman et al. 2006; Friedman and Hendry 2019). Friedman developed the approach of 'value-sensitive design,' an approach to account for and incorporate human values in a comprehensive manner throughout the design process. This approach was initially developed for the design of information systems but is more broadly applicable. It proposes investigations into values, designs, contexts of use, and stakeholders with the aim of designing systems that incorporate and balance the values of different stakeholders. The key activities in value-sensitive design are the identification of direct and indirect stakeholders and the benefits and harms for each group that may result from the system that is to be designed (empirical investigations); the mapping of benefits and harms onto corresponding values; conceptual investigations of key values and the identification of potential value conflicts and the proposal of solutions for them (conceptual investigations); and studies of how properties of the to-be-designed artifact may support or counteract human values and the artifact may be designed proactively in order to support specific values that have been found important in the conceptual investigation.

Many scholars have been inspired by the value-sensitive design approach, and while some work within its scope, others have developed alternative approaches for incorporating values into design. The term 'design for values' is sometimes used to denote the broader family of design approaches that incorporates the idea of value embeddedness (van den Hoven et al. 2015b). As I have argued (Brey 2010b), different approaches to design for values hold different positions on how the relevant set of to-be-promoted values should be identified (e.g., through stakeholder consultation, normative ethical analysis, consultation of constitutions and (inter) national declarations on rights and ethics, or combinations thereof); how value conflicts should be resolved (through deliberation by stakeholders, consultation of stakeholders, normative analysis, or other means); and how values can be translated into design requirements.

It is important to realize that design for values approaches are not necessarily constrained to moral values. They are sometimes thought of as such, and there are a few design-for-values approaches that have a more specific focus on morality and ethics. However, most approaches, including value-sensitive design, consider non-moral values as well. Values come in many sorts, and next to moral values, one can find, among others, aesthetic, economic, social, cultural, epistemic, spiritual, and personal values. As I argued in section 3.1, moral values are important to society, as they allow one to distinguish right from wrong, but they do not define the totality of what is valuable or good. Therefore, as I argued, a good society is not the same as one in which moral values are realized. However, a broader design for values approach that includes non-moral values as well could be a viable approach for design for a good society, because a good society can, at least to a considerable extent, be defined in terms of a set of values that should be realized for a society to be good. If one is only interested in ethical design, then design for values approaches are also of use; one simply makes the choice to only consider moral values in the value selection process.

A shortcoming of values in design approaches is that they do not include detailed and rigorous design methodologies that specify how conceptual, empirical and technical investigations should proceed and should be integrated with each other (Manders-Huits 2011). There is often no detailed methodology for identifying and surveying stakeholders, for translating stakeholder benefits and harms to values, for making value trade-offs, for translating values into design requirements, and for integrating design for values approaches with "mainstream" design methodologies. Recent work attempts to address some of these issues within value-sensitive design (Friedman et al. 2018) and in other approaches (van de Poel 2015; Kroes and van de Poel 2015; Vermaas et al. 2015).

In the remainder of this section, I will make a modest contribution to this recent development by considering how design for values approaches can be related to the account of design processes by Pahl et al. that was discussed in section 2.3. The most important phase in the Pahl et al. account to incorporate value issues is, I claim, the task specification phase, which comes after the initial product planning phase. In the task specification phase, the kind of product that is needed is clarified, and requirements and constraints are identified and formulated, resulting in a requirement list. Naturally, during this phase, values would be identified and included in the requirement list. For example, at this point it could be specified that the product should protect the privacy of users and other stakeholders, or that it should be supportive of the overall well-being of users. At this phase, recommendations and requirements regarding value trade-offs could also be made. This is not to say that values should not be considered at all during the prior product planning phase. If values are front and center during this phase already, then it is less likely that product ideas will be developed that are incompatible with relevant values, and that later discovery of this fact, if it takes place at all, requires a radical redesign.

During the subsequent conceptual design phase, conceptual-level design solutions are found for the challenge posed at the task clarification stage. For example, the design of an information system would include a conceptual specification of basic functions and subfunctions of the system, and working principles for these subfunctions and their combination into working structures. If one of the design specifications is that the system should be protective of the privacy of the users, then at this phase, design solutions are sought in which no personal information from the user is recorded, such recording is by design temporary, or such recordings are contained so that they are not accessible by third parties. For some values, the level of abstraction of the conceptual design phase may be too high to enable specifying design features that are relevant for their realization, and these values could come into focus later, at the embodiment design or detail design phases. What is needed, and does not exist at this point, is a general methodology for operationalizing and integrating value requirements at the conceptual design phase, including conceptual-level operationalization of particular values.

Next, at the embodiment design phase, the product is defined at a more concrete level, including its general arrangement, spatial features, shape and materials, and auxiliary functions not covered at the conceptual design stage. At this stage, concrete implementations need to be found for the conceptual-level solutions found for the inclusion of values in the conceptual design phase. For example, if during that phase, it was found that personal information input by users should only be stored temporarily, now a specific solution is needed for how this is done, for example by only storing such information in a dedicated of section of RAM and having algorithms in place that prevent it from being stored permanently. Also missing at this point for this stage in the design process are general as well as value-specific methodologies for translating conceptual-level value solutions to embodiment-level solutions.[6] Finally, the embedding of values may also partially take place during the detail design phase, when a definitive determination of shapes, properties, materials, and production methods is made. Because the success of designing for values is to be measured by the success a design has in actually promoting these values when in actual use, testing, including consumer testing (or better: stakeholder testing) could also be an important component of value in design approaches, as would evaluation and possible redesign based on investigations of market response and possibly also social and ethical impact assessments that are performed after introduction to market.

4. Conclusion

In this chapter, I investigated two issues: the nature of engineering design and the moral, social, and political choices embedded in design. Engineering design was related to other types of design and other human practices, and was defined as the development, through the application of science, mathematics, and engineering knowledge, of plans for products (devices, systems, methods, procedures) that can serve practical ends. It was argued, as well, that engineering design is distinguished from other (design) practices by its unique methods, produced knowledge, and manufactured products. Engineering design was also situated among other engineering practices, and was related to technological innovation, in which innovative design often, but not always, plays a significant role. I also considered the structure of the design process, and examined an influential conception of it by Pahl et al. (2007), which distinguishes five phases in design.

I then turned to the ethical, social, and political dimensions of design. I started by distinguishing different meanings of "good design" and analyzing the relation between engineering design, on the one hand, and values, benefits, and the good of society, on the other. I argued that the most important type of goodness of a technological product is its goodness for society, and that other types of goodness (functional goodness, specifications goodness, moral goodness, prudential goodness for aspects of society) are subordinate to it. I then investigated how values and consequences for society can be embedded in design, a theory of which is needed for formulating approaches to design that are beneficial to society and its constituent elements. Finally, I investigated and critiqued approaches for designing for values and benefits for society, and made my own contributions to this debate.

Approaches to design that focus on values and benefits to society have a lot of promise, but methodologies for them need to be developed more and integrated with mainstream design methodologies. If this were to happen, they could eventually become part of the mainstream engineering education. There is certainly a lot of interest in society in the development of technology that is ethical and beneficial to society. It should be considered, though, that these approaches may be best applied by multidisciplinary teams, which include members with training in humanities and social sciences, or engineering teams in which some of the engineers have a multidisciplinary background. The take-up of this kind of approach ultimately depends on the interest of commercial firms in developing technologies in this manner, taking into account that technology development takes place for the most part in the private sector. It will depend on the way firms conceive of and implement corporate social responsibility, and on the legislation and regulations that will be in place to constrain and guide design and manufacture.

Notes

[1] The term "social design" is also used to refer to engineering design activities aimed at solving social problems.

2. Instrumental goodness is akin to what Von Wright (1965) has called instrumental goodness: the goodness of instruments or tools of type *X* as type-*X* instruments. For example, if a drill (or other object) is good as a drill (i.e., performs the drilling function well), then it has instrumental goodness as a drill. See also Ylirisku and Arvola (2018), who distinguish various meanings of the term 'good design.'
3. While 'prudential goodness' or value is usually attributed to persons and relates to their well-being, I use it here in a broader sense, to denote value that can benefit (contribute to the flourishing of) any kind of thing that can be benefited.
4. This thesis refers to the impact neutrality of technological products. There is also a neutrality thesis that refers to value neutrality or moral neutrality, e.g., Morrow (2014).
5. The approach of Responsible Research and Innovation (Von Schomberg 2013) is also directed at ensuring technological innovations that make a better fit with society and provide more social benefits. It is an overall strategy toward the research and innovation system that includes design as only one element.
6. For a few values, such methodologies have been developed to some extent, both for the conceptual and embodiment design phase. For example, in the approach of privacy by design (Cavoukian 2012).

References

ABET. 2018. Criteria for Accrediting Engineering Programs, 2018–2019. Accessed June 14, 2021. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2018-2019/.

Bardzell, Jeffrey, Shaowen Bardzell, and Mark Blythe, eds. 2018. *Critical Theory and Interaction Design*. Cambridge, MA; London, England: MIT Press.

Boon, M. 2011. "In Defense of Engineering Sciences." *Techné* 15 (1): 49–71.

Brey, Philip. 2000. "Disclosive Computer Ethics: The Exposure and Evaluation of Embedded Normativity in Computer Technology." *Computers and Society* 30 (4): 10–16.

Brey, Philip. 2005. "Artifacts as Social Agents." In *Inside the Politics of Technology. Agency and Normativity in the Co-Production of Technology and Society* , edited by Hans Harbers, 61–84. Amsterdam: Amsterdam University Press.

Brey, Philip. 2006. "The Social Agency of Technological Artifacts." In *User Behavior and Technology Development: Shaping Sustainable Relations Between Consumers and Technology*, edited by Peter-Paul Verbeek and Adriaan Slob, 71–80. Dordrecht: Springer Netherlands.

Brey, Philip. 2010a. "Philosophy of Technology after the Empirical Turn." *Techné: Research in Philosophy and Technology* 14 (1): 1–13.

Brey, Philip. 2010b. "Values in Technology and Disclosive Computer Ethics." In *The Cambridge Handbook of Information and Computer Ethics*, edited by Luciano Floridi, 41–58. Cambridge: Cambridge University Press.

Brey, Philip. 2018. "The Strategic Role of Technology in a Good Society." *Technology in Society* 52: 39–45.

Cavoukian, Ann. 2012. "Privacy by Design: Origins, Meaning, and Prospects for Assuring Privacy and Trust in the Information Era." In *Protection Measures and Technologies in Business Organizations: Aspects and Standards*, edited by George Yee, 170–208. Hershey, PA: IGI Global.

Chakrabarti, Amaresh, and Lucienne T. M. Blessing, eds. 2014. *An Anthology of Theories and Models of Design: Philosophy, Approaches, and Empirical Explorations*. London: Springer-Verlag.

Dym, Clive. 1994. *Engineering Design: A Synthesis of Views*. New York: Cambridge University Press.

Eppinger, Steven D., and Greg Geracie. 2013. *The Guide to the Product Management and Marketing Body of Knowledge*. Reno, NV: Product Management Educational Institute (PMEI).

Flanagan, Mary, Daniel C. Howe, and Helen Nissenbaum. 2005. "Values at Play: Design Tradeoffs in Socially-Oriented Game Design." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 751–760. CHI '05. New York, NY: ACM.

Fletcher, Guy. 2012. "Resisting Buck-Passing Accounts of Prudential Value." *Philosophical Studies* 157 (1): 77–91.

Fogg, B. J. 2002. *Persuasive Technology: Using Computers to Change What We Think and Do*. 1st ed. Amsterdam, Boston: Morgan Kaufmann.

Friedman, Batya, and David G. Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press.

Friedman, Batya, David G. Hendry, and Alan Borning. 2018. *A Survey of Value Sensitive Design Methods*. Boston Delft: Now Publishers Inc.

Friedman, Batya, Peter Kahn, and Alan Borning. 2006. "Value Sensitive Design and Information Systems." In *Human-Computer Interaction in Management Information Systems: Foundations*, edited by P. Zhang and D. Galletta, 348–372. Armonk, NY: M.E. Sharpe.

Gero, John S. 1990. "Design Prototypes: A Knowledge Representation Schema for Design." *AI Magazine* 11 (4): 26.

Hoven, Jeroen van den, Pieter E. Vermaas, and Ibo van de Poel, eds. 2015a. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values, and Application Domains*. Dordrecht: Springer Netherlands.

Hoven, Jeroen van den, Pieter E. Vermaas, and Ibo van de Poel. 2015b. "Design for Values: An Introduction." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values, and Application Domains*, edited by Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel, 1–7. Dordrecht: Springer Netherlands.

Jack, Hugh. 2013. *Engineering Design, Planning, and Management*. London: Academic Press.

Johannesson, Paul, and Erik Perjons. 2014. *An Introduction to Design Science*. Cham: Springer International Publishing.

Kroes, Peter. 2012. *Technical Artefacts: Creations of Mind and Matter: A Philosophy of Engineering Design*. Dordrecht: Springer Science & Business Media.

Kroes, Peter, and Meijers, Anthonie, eds. 2001. *The Empirical Turn in the Philosophy of Technology*. Research in Philosophy and Technology 20. London: Elsevier/JAI Press.

Kroes, Peter, and Ibo van de Poel. 2015. "Design for Values and the Definition, Specification, and Operationalization of Values." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, edited by Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel, 151–178. Dordrecht: Springer Netherlands.

Latour, Bruno. 1990. "Technology Is Society Made Durable." *The Sociological Review* 38 (1): 103–131.

Malerba, Franco, and Luigi Orsenigo. 1997. "Technological Regimes and Sectoral Patterns of Innovative Activities." *Industrial and Corporate Change* 6 (1): 83–118.

Manders-Huits. 2011. What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics* 17 (2): 271–287.

Meijers, Anthonie. 2009. *Philosophy of Technology and Engineering Sciences*. Vol. 9 of *Handbooks of the Philosophy of Science*. Series editors Meijers, Anthonie, Dov Gabbay, Paul Thagard, and John Woods. Burlington, MA: Elsevier.

Morrow, David R. 2014. "When Technologies Make Good People Do Bad Things: Another Argument against the Value-Neutrality of Technologies." *Science and Engineering Ethics* 20 (2): 329–343.

Nissenbaum, Helen. 1998. "Values in the Design of Computer Systems." *Computers and Society* 28 (1): 38–39.

OECD/Eurostat. 2018. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation*. 4th ed. The Measurement of Scientific, Technological, and Innovation Activities. Paris/Eurostat, Luxembourg: OECD Publishing.

Pahl, Gerhard, and W. Beitz. 1996. "Engineering Design: A Systematic Approach." *MRS Bulletin* 21 (8): 71.

Pahl, Gerhard, W. Beitz, Jörg Feldhusen, and Karl-Heinrich Grote. 2007. *Engineering Design: A Systematic Approach*. 3rd ed. London: Springer-Verlag.

Parsons, Glenn. 2015. *The Philosophy of Design*. 1st ed. Cambridge, UK: Polity.

Poel, Ibo van de. 2015. "Design for Values in Engineering." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, edited by Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel, 667–690. Dordrecht: Springer Netherlands.

Rose, Ellen. 2012. "Not 'Just a Tool': A Triadic Model of Technological Non-Neutrality." *Educational Technology* 52 (1): 17–21.

Sachs, Angeli, Claudia Banz, and Michael Krohn. 2018. *Social Design*. Zürich: Lars Müller Publishers/Museum für Gestaltung Zürich.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.

Verbeek, Peter-Paul 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park: Pennsylvania State University Press.

Verbeek, Peter-Paul. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press.

Vermaas, Pieter E., Paul Hekkert, Noëmi Manders-Huits, and Nynke Tromp. 2015. "Design Methods in Design for Values." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, edited by Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel, 1–19. Dordrecht: Springer Netherlands.

Vermaas, Pieter E., Peter Kroes, Andrew Light, and Steven Moore, eds. 2008. *Philosophy and Design: From Engineering to Architecture*. Dordrecht: Springer Netherlands.

Vermaas, Pieter E., and Stéphane Vial, eds. 2018. *Advancements in the Philosophy of Design: Design Research Foundations*. Cham: Springer International Publishing.

Vincenti, Walter. 1990. *What Engineers Know and How They Know It*. Baltimore, MD: Johns Hopkins University Press.

Von Schomberg, René. 2013. "A Vision of Responsible Research and Innovation." In *Responsible Innovation. Managing the responsible emergence of Science and Innovation in Society*, edited by Richard Owen, John Bessant, and Maggy Heintz, 51–74. West Sussex: Wiley.

Von Wright, Georg Henrik. 1965. *The Varieties of Goodness*. London: Routledge & Kegan Paul.

Wendel, Stephen. 2013. *Designing for Behavior Change: Applying Psychology and Behavioral Economics*. 1st ed. Sebastopol, CA: O'Reilly Media.

Willis, Anne-Marie, ed. 2018. *The Design Philosophy Reader*. Annotated ed. London: Bloomsbury Visual Arts.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109 (1): 121–136.

Xu, Zhenning, Gary L. Frankwick, and Edward Ramirez. 2016. "Effects of Big Data Analytics and Traditional Marketing Analytics on New Product Success: A Knowledge Fusion Perspective." *Journal of Business Research*, Designing implementable innovative realities 69 (5): 1562–1566.

Ylirisku, Salu, and Mattias Arvola. 2018. "The Varieties of Good Design." In *Advancements in the Philosophy of Design*, edited by Pieter E. Vermaas and Stéphane Vial, 51–70. Cham: Springer.

CHAPTER 21

VIRTUAL REALITY MEDIA AND AESTHETICS

GRANT TAVINOR

1. Technology and the Arts

Technologies are in general amplifications of our natural powers and technological innovations have always had an impact on the arts and artistic practice, often by providing new means of artistic expression. To take an obvious example, photography quickly found aesthetic applications upon its development in the nineteenth century. The photographic movement of pictorialism—where the photographs were altered, manipulated or combined to bring attention to the surface or expressive qualities of the form—was developed by artists such as Henry Peach Robinson and Alfred Stieglitz, who then refined and expounded the artistic principles of the movement and argued for the status of such photography as an art form. Advancing a different conception that emphasized the medium's ability to produce crisp and detailed images, artists such as Ansel Adams produced works that emphasized the abstract forms and textural qualities of the image, in a way that they thought employed the distinctive capacities of photography as an image-making technology.

This artistic application of photography led to the investigation of the theoretical and philosophical issues inspired by the new art form, and it raised many philosophical questions: How does photography differ from previous ways of image making such as drawing or painting? Is photography a matter of art or mere documentation? Does the apparent "mechanical" nature of photography affect the expressive capacity or ontology of the artistic medium of photography? Are photographs a uniquely realistic—moreover—truthful or transparent medium? Thus, in addition to the substantive developments in artistic media, technological developments such as photography clearly have theoretical consequences for our understanding of the artistic expression, ontology, and criticism of the aesthetic practices and art forms they produce. Roger Scruton (1981), Kendall Walton (1984) and Greg Currie (1999) have made key contributions to the philosophical issues prompted by the artistic use of photography.

Virtual reality (VR) media are perhaps the most recent of such aesthetically and artistically fruitful technological developments. The concept of virtual reality has been with us for at least 40 years and has had an undeniable influence on popular culture—particularly as a subject of traditional media in films such as *Tron* (1982) or *The Matrix* (1998)—but it is only recently that the medium itself has become widely available for home use. Several VR products are now commercially available, including the PlayStation 4 VR and the HTC Vive, tethered and standalone headsets from the VR forerunner Oculus, and a number of mobile headsets that use smartphones as their screen, such as Samsung Gear VR and the Google Daydream View.

While VR headsets have found aesthetic applications in the fine arts, it is in the popular phenomenon of videogames that the real impact of virtual media can best be seen, and they will be my focus here. A good example is *Resident Evil VII*, a survival horror game that utilizes VR media to situate the player within a deserted, poorly lit, and decaying house. The game also includes insane cannibals. *Resident Evil VII* can be a terrifying experience, largely because of the vulnerability that results from being situated within the virtual world of the game. In fact, such was the sense of personal fear I felt while playing that I had trouble finishing the game. Such games are far from perfect—VR sickness is still a significant issue—and it is not yet clear how lasting the current interest in VR gaming will be, but *Resident Evil* and other games do adopt the new medium to a frequently striking effect.

Virtual reality media are of intrinsic philosophical interest, but the ongoing development of the medium is also significant because of how its consideration casts light on some traditional aesthetic concerns, including the nature of depiction, artistic interpretation, the concept of fiction, and the status of an appreciator's emotional responses to artworks. In this chapter I will focus on the first of these issues and provide an analysis of VR depictive media framed against the historical development of perspectival depiction.

The next section involves a brief discussion of some of the technological developments in VR media. Because of the astounding rate of development of the technology, this discussion can only be very incomplete: however, I pick out several factors that are crucial to understanding the potential contribution of VR to artistic expression. Part three of this chapter investigates the definition of virtuality itself, a concept that is frequently quite ambiguous in application. I evaluate what can be said about the nature of virtual reality technology given a more precise analysis of virtuality. A key part of virtual media thus analyzed, is their depiction of an apparent appreciative viewpoint within a virtual world. Section 4 of this chapter links VR media to the wider consideration of perspectival depiction in artistic media. The development of linear perspective in the fifteenth century is sometimes treated as an attempt by artists such as Brunelleschi to replicate natural human vision. Careful analysis of the techniques of linear perspective shows this ambition to be only roughly achieved within the artistic traditions that employed it. VR media, however, seem to improve on several of the apparent failings in earlier modes of perspectival depiction. In the final part of the chapter I explore one way in which virtual reality media are exploited to artistic effect: the self-involvement of virtual media in works such as *Resident Evil VII* gives the player a deepened sense of vulnerability and may result in distinctive emotional responses. I explain how the apparent appreciative viewpoint key to virtual self-involvement is precedented in previous art, and how these depictive precedents can be used to inform our understanding of self-involvement in VR media.

2. Virtual Media Technology

In the most general terms, PlayStation VR, the HTC Vive and other commercial VR systems comprise three key elements: the depiction of a sensory environment; a means of tracking and depicting the user's apparent position within this environment; and, finally, a means of interacting with this virtually depicted space.

The depiction of the VR environment is principally visual and is most frequently achieved via a stereoscopic headset. For example, the PlayStation VR headset includes a single 5.7-inch organic light-emitting diode (OLED) panel with 1080p display resolution and a potential refresh rate of 120 frames per second. The resolution and refresh rate—which contribute to the smoothness of the displayed movements—are important if the depicted world and its objects are to give the impression of solidity and permanence. Two small lenses are placed before the screen, magnifying and softening the images, allowing for a wider field of view on the depicted images, and reducing the perceptual prominence of the surface of the pixel array. The images that are depicted on the panel are produced in such a way—and this is where the sophisticated rendering algorithms of the software play a role—that the apparent displacement of the two images, combined with the binocularity of the lenses, mimics our natural visual situation in the real world. The resulting visual environment—such as a darkened kitchen within a decaying house—gives a very strong visual impression that the viewer is situated within that environment.

The depiction of the sensory environment is not restricted to this visual modality, and at a minimum VR media usually involve stereophonic sound to place the user within an aural space. The spatial effect is achieved in a formally identical manner to the stereoscopic depiction of the visual environment: the spatial cues of native hearing—the displacement of the two ears and the brain's ability to use the resulting difference in timing and intensity of the received sounds to identify the spatial location of sound sources, that is, the ability of *sound localization*—is utilized in VR to mimic the acoustic spatiality of natural hearing. But it should be noted that this technique is also used in what we would not ordinarily consider VR media, such as where traditional cinema employs stereo sound to give objects and events of the diegetic world a spatial location in relation to the film viewer. A good example is Steven Spielberg's *Saving Private Ryan* where the approaching rumbling of a Tiger tank plays a key dramatic role in one sequence. The formal equivalence of virtual stereoscopic vision and cinematic stereo sound may give us pause to attributing a theoretical novelty to the recent developments in VR media, a general issue I will take up later in this chapter.

Sometimes a means of haptic or kinesthetic depiction is also used to convey a sensory engagement with the virtual environment; the senses in question being touch, proprioception and the spatial sense provided by the vestibular system. However, these modes of sensory representation are less common and quite limited compared to the visual and auditory modes already discussed.

This apparent perceptual vantage point is not passive or fixed, as VR systems usually allow for the movement of the user to be tracked and for this movement to be represented in the user's apparent perceptual orientation on the virtual environment. In PlayStation VR, this tracking involves 9 small LEDs fixed to the exterior of the headset which are captured by a camera placed in front of the user. The relative position of the LEDs is used to calculate the orientation of the user's head so that head movement can be replicated in the depictive viewpoint displayed by the stereoscopic headset. The HTC Vive goes further and allows for the user's bodily movements to be tracked and depicted within a 15-foot radius. Other VR systems such as Oculus and Magic Leap have developed eye tracking technology so that the perceptual effects of eye movement—particularly on focus in the visual field—might be replicated. For each of these means of tracking to be effective, a low latency between the tracking and the display is crucial for creating a realistic impression; this is a significant technical hurdle in producing effective VR systems.

While many VR media applications comprise "experiences" where the user's agency is limited to changing their orientation on the depicted virtual spaces, most applications, including all VR games, allow the user the ability to interact with objects within the virtual space. So, for example, in the PlayStation VR game *The London Heist* (2016) the player may interact with objects such as cigarette lighters and mobile phones: in one scene the player lights the cigar of another character. Such control is typically achieved either through standard gaming controllers or purpose-built peripherals, but again the position and orientation of these devices needs to be tracked by the VR system so that the user's movements can be replicated in the virtual space. An important variation in controls is between control movements that are not themselves depicted in the virtual space—such as pressing a gamepad button to perform the virtual action of opening a door—and control movements that are given virtual representations, such as where one might aim a prop gun to aim a gun in the virtual world, or lift a controller to their ear to virtually lift a mobile phone to their ear (Tavinor 2018, 158). The latter kinds of control are referred to as gestural controls. I will not have much to say about the means of interaction offered by virtual media, though a full account of VR media and art would do just that; rather, my focus will be on VR as a technological expansion of the depictive means of art.

3. What Are VR Media?

In what sense are such technological developments all properly referred to as progresses in *virtual* media? In the wide and diverse literature on virtual reality, and in its everyday discussion, the use of the term "virtual" to refer to virtual worlds, virtual stores and currency, virtual media, virtual computers and memory, or to the basic concept of virtuality, is often extraordinarily vague, so much so that one might have doubts about the real utility of the term. Like the term "interactivity," one might easily suspect "virtuality" of being a mere technological buzzword, referring quite imprecisely to a range of phenomena united only by their technological setting. Furthermore, virtual reality theorist Michael Heim suspects that we are likely to simply become confused by the term if we attempt to analyze it in terms of apparent symptoms such as "immersion," "telepresence" and "interaction" (Heim 1993, 109ff).

There is, however, a reasonable way to refine the intension of the term to see its genuine theoretical utility (though as we will find later this refinement may also result in a reconstruction of the appropriate extension of the term so that it comes to include some objects or activities we might not have previously considered to be "virtual"). This intentional refinement can be achieved by taking seriously a core usage of the word "virtual," one defined by Charles Sanders Peirce where, "A virtual X (where X is a common noun) is something, not an X, which has the efficiency (virtus) of an X" (Peirce 1935, 261). Non-technically, a virtual store such as Amazon is *as good as* a brick and mortar store for the purpose for which we find stores useful: Amazon has "the efficiency of" a brick and mortar store because it achieves the same function. In a more technical sense, I have argued elsewhere that useful in understanding this sense of virtuality is the idea of "structural or functional isomorphism" (2018, 154). Isomorphism is a term with applications in biology, crystallography, and mathematics that means "equal form" and is used to refer to a functional or structural correspondence between objects in different material domains.

Given the as-if sense of virtuality previously identified, what is a virtual reality medium? There is of course a more fundamental question to clear up: What is a *medium* anyway? A rather unsatisfactory answer to this question would involve reference to *the media*—that is, the prevalent means of mass communication—and to see virtual media as a mere digital or computational addition to or extension of these. But *medium* is a concept that is key to our understanding of a wide range of events and processes beyond this narrow communications context and acknowledging this wider usage does throw light on virtual media. A basic physicalist sense of the term is of an intervening substance or activity through which physical forces or movement is transmitted. A mentalistic parallel of this is of a substance or activity via which sensory impressions or experiences are conveyed. Finally, we might consider as a communicative medium one through which ideas or information are conveyed between individuals. The experiential sense seems to be an obvious focus in the current context given that it is the presentation of sensory environments—*virtual spaces or realities*—that is the most striking aspect of virtual media. But virtual reality media are not limited to providing experiences. An example we shall look at presently illustrates how such media can be the means of achieving physical movement or performing actions. And, of course, VR media will often comprise a communicative medium where their virtual realities allow for communication between individuals.

Whichever of these three senses is the focus, given that the focus of this chapter is virtual media and aesthetics, there is a further immediate question of the relationship of VR media to *artistic media*.[1] While they are capable of conveying or communicating movement, impressions, experiences or ideas, we need not assume that virtual media necessarily do so with any aesthetic or artistic intent or consequence. Non-artistic uses of VR media exist within education, medicine, and military contexts. But where VR media are employed in the context of artistic expression or communication, their status and potential as an artistic medium is likely to prove to be an important concern. The final part of this chapter takes up this issue. However, though I will focus on the virtual replication of native perceptual experience, this is not the only way that virtual media may embody artistic media: virtual dance and virtual painting are other potential uses.

My focus here, then, is the sense in which VR media convey the features we ordinarily associate with natural perceptual experience, allowing users sensory access to an apparent virtual world; moreover, the way in which virtual media frequently bear a striking functional and structural correspondence with ordinary experience. A recent set of experiments can help explain the issues here. In Mathew Pan and Günter Niemeyer's work at Disney Research (2017), a user wearing a stereoscopic headset is able to catch a real ball that is tossed to them, based entirely on the motion-tracked ball's depicted movement in a VR environment.[2] Employing a motion-tracking camera, a ball's position is tracked and displayed on the stereoscopic headset as an animated ball within a rudimentary virtual environment comprising a textured floor, basic lighting, and paddle-like depictions of the user's hands (Pan and Niemeyer 2017, 1). Using this visual information, the user can orientate her hands to achieve the task of catching the ball in virtual and in real space.

In this research the virtual medium, employing a stereoscopic headset and spatial tracking, has the efficiency of a real perceptual engagement in such a situation; indeed, in this case it *replicates* native perceptual engagement, counting as a kind of prosthetic seeing (Tavinor 2019a). Like other technological innovations, the depictive medium counts as a potential amplification of our natural powers or abilities. Furthermore, here the analysis of virtuality as structural and functional isomorphism is relatively clear: because the spatial structure of the trajectory of the real ball is replicated within the apparent space of the stereoscopic depiction, the medium has the functional efficiency of native vision, at least for the physical task of catching the ball.

One could also extend the analysis of virtuality offered here to account for the meaning of phrases such as "virtual world": a virtual world would be the apparent experiential space depicted in the medium, and that allows for this perceptual and functional engagement seen in the Disney research. A confusion that is looming, however, and one that I have purposefully resisted in my framing of these issues, is that talk of virtual worlds and objects easily falls into metaphysical speculation. Many philosophical accounts of VR begin with Cartesian considerations of "perfect matrix-like" virtual realities (Chalmers 2003) or of the metaphysics of virtual worlds (Heim 1993) and then quickly dive down an ontological rabbit hole. One focus of such accounts is the contested relationship between virtual worlds and fictional worlds, with some philosophers resisting the idea that virtual worlds are fictional worlds (Chalmers 2003, 2017). However, an important thing to note about the Disney research is that what is depicted by the VR medium is not a fiction, such as a deserted and decaying house, but a real ball and its real trajectory through space. The Disney experiment allows for a virtual engagement with the real world, but an arguably more frequent use of the technology is where it is used to allow for apparent interaction with fictional worlds such as in *Resident Evil VII*.

At the very least, beginning with a metaphysical orientation gives the resulting discussions a very different flavor; but I also believe that it is preferable to engage with this topic principally under the rubric of *virtual media* and to focus on how we currently use VR media because this will allow us to avoid the abundant confusions that a metaphysical approach encourages. One prevalent confusion regards the status of what users "see" when they don headsets, with some virtual "realists" holding the implausible idea that users do not see the screen, but rather see virtual items directly (Chalmers 2017). Orientating instead on media, it should be clear that users *do* see the screen—and sometimes, because of artifacts such as the so-called "screen door effect" caused by the pixelation of the screen, it is particularly prominent to the user—even if they see depicted objects *through* this screen (Tavinor 2019a).[3] Even more strongly, it could be best not to speak of virtual worlds or realities at all: rather, we might restrict ourselves to talk of virtual media, which are capable of depicting fictions or reality. For this reason, elsewhere I have developed the distinction between "virtual realism" and "virtual fictionalism," a distinction that depends on the intentional context of uses of the medium (Tavinor 2019a). Virtual realism involves the mediation, through virtual technologies such as stereoscopic headsets, of a perceptual and causal interaction with the real world. Virtual fictionalism is intended to mediate, through the same technologies, a *fictional* causal and perceptual interaction with an imaginary world.

Furthermore, by employing a reasonably developed theory of fiction, we can understand how virtual media can depict fictions without courting any metaphysical mystery: virtual media often count as what Kendall Walton would call "props" for the imagination, in that they make things fictional of a game of make-believe (Walton 1990). To conclude this section, the concept of virtual media is preferable to virtual worlds or realities because it serves to deflate the frequent metaphysical issues that threaten our understanding of these technological developments.

4. Virtual Media and Perspectival Depiction

I have claimed that VR media give—or are intended to give—their users a perceptual engagement with a depicted space that is isomorphic with native perceptual access to the world. This is not the first time such attempts have been made, as the development of linear perspective in the fifteenth century is sometimes considered as such an attempt to replicate native vision. The crucial developments in this regard were experiments performed by Filippo Brunelleschi in Florence in 1413 (Edgerton 2009). Brunelleschi painted a panel depicting the Florentine Baptistery in his newly developed linear perspective style. By holding the panel with the image facing away from him toward the Baptistery, through a hole in the panel the building could be compared with the image by inspecting the painting with a mirror. Brunelleschi's intention was to directly compare his painting technique with the qualities of natural vision to show how his technique rendered the geometry of natural vision in a potentially illusionistic way.

Leon Battista Alberti went beyond these experiments in his codification of linear perspective in the treatise *De pictura* (1435). Alberti designed a system of perspective—a *perspectiva artificialis*—that treated painting in terms of an imaginary plane suspended between the viewer and the depicted scene. This plane acts as the base of a vertically tilted pyramid, where the peak of the pyramid extending away from the viewer represents the vanishing point of light rays. Employing a grid that is drawn on both axes of the picture plane, a horizon line that runs horizontally through the vanishing point, and orthogonal lines that run perpendicular to the picture plane, the bottom side of the tilted pyramid can be depicted as a foreshortened tiled floor. This is called the "plan and elevation" method of perspective (*costruzione legitimma*) and its use allows for figures and objects to be appropriately scaled and positioned in the picture space. Alberti's system was an important technological advancement because the perspectival techniques used by artists stretching at least as far back as the Romans were intuitive, accidental or less than theoretically informed, in comparison to the coherent theory seen in Alberti's plan and elevation method. Alberti's technological contribution for conveying convincing picture space was applied by contemporary artists and was powerfully influential on the future artistic depiction of space (Hyman 2006, 217-223), famously in Pietro Perugino's fresco in the Sistine Chapel, *Christ Giving the Keys to St. Peter* (1481–1482).

While it is often impressive in application, even in some cases generating a kind of illusionistic effect, linear perspective does not exactly replicate the features of native human visual perception. First, the perceptual results of binocularity are not reproduced by linear perspective: in the experiments made by Brunelleschi, because there is only one image, a single set of rays converge at the vanishing point and the depiction remains flat in comparison to the depth of binocular vision. Linear perspective is also constrained by the fixed nature of such depictions: in native vision the eyes and head are always moving, and the static nature of painting cannot reproduce this movement and its effects on the apparent spatiality of the scene. Thus, if the viewer of a painting produced under Alberti's pictorial conventions does have a precisely defined vantage point of those scenes, it is an unnaturally fixed one, and usually one that is fixed in a specific position quite close to the picture surface. Furthermore, linear perspective leads to distortions, particularly at the edges of images where objects become unnaturally stretched. Linear perspective also depicts a limited field of vision compared to natural vision (partly this is necessary to avoid the spatial aberrations resulting from wide-angle linear perspective depictions). Finally, there is even experimental evidence that linear perspective is not judged as a realistic depiction of space compared to alternative modes of depiction (Burleigh, Pepperell, and Ruta 2018). Thus, linear perspective remains only a partial and imperfect approximation of native vision.

These problems discount the idea that there can be purely geometrical schemes of spatial depiction that match native vision, given that geometrical idealizations are likely to render unconvincing scenes. Furthermore, geometrical perspective would not by itself be sufficient for the rendering of convincing visual scenes: in native vision, occlusion, light and shade, and textural gradients all add additional cues to the spatial depth of visual scenes, and these cues have frequently been incorporated in traditional art (Kubovy 1986). Foregoing strict linear perspective, Jan van Eyck's *Arnolfini Portrait* employs the light source provided by a window to give an effective sense of volume to the represented space; the brightly lit space beyond the two standing figures contrasts with some of the darker foreground areas. Another spatial cue in large landscape scenes is the diffusion of light sources through the atmosphere, and the bluish washed-out appearance this gives to distant objects. This shows that depictive naturalism is not merely a matter of rendering spatial geometry, but that it also relies on providing the kinds of visual cues that our native vision employs to register the spatial depth of scenes. The cognitive psychology of vision has now investigated the kinds of cues critical to spatial perception and their use in art (Kubovy 1986).

VR media has drawn on many of the achievements of the representation of space in traditional art—including the principles of linear perspective—but it has also sought to improve on some of the weaknesses of earlier techniques in spatial depiction. VR media can thus be understood as a technological advancement of these modes of perspectival depiction, meeting some of these limitations of traditional static depictions.

A first key development here is the allowance for the free movement of the apparent point of view of the viewer within the depictive space. This development predates current VR and owes to the development of 3D graphical environments and the virtual camera (Kerlow 2000, 88–91). In videogames, the development of 3D graphics has been an essential element in expanding the range and appeal of the form. The depictive surfaces of most videogames are ultimately the 2D picture planes—whether on television screens or personal computer displays—that might be usefully analyzed under Alberti's method of plan and elevation (with the complication that their 2D picture spaces are animated). But the production of these 2D pixel arrays employs a very different technique: that of texture-mapped wireframe objects—essentially mathematically defined geometrical models—via which the objects and environments of the depicted world are built up, and a "virtual camera" that is used to define the perspective on the digital environment that is rendered on the 2D screen. The orientation of the virtual camera is frequently identified with that of the player-character in the world, though in third-person games the camera may float above and behind the character; but in both cases the virtual camera gives the player a view into the game world.

Moreover, the player is usually in control of this depictive perspective and can move through the depicted world. Early 3D games such as the proto-first person shooter *Wolfenstein 3D* allowed for movement in two dimensions through very simple spatial environments (in *Wolfenstein 3D*, a series of uniform rectilinear hallways and rooms). Playing the game involves exploring this space, defeating enemies (Nazis) and looking for an elevator that leads to further areas. Hence, unlike traditional painting where the apparent viewer's orientation on the depictive space is fixed and the whole of the picture space is present on the picture plane in the moment portrayed in the painting, in modern 3D graphics the depictive space is extended spatially, and what is actually rendered at any one time counts as only a portion of what is potentially portrayed of the world. The visual scenes reproduced in 3D graphics are no longer static with respect to the position of the viewer, as through the manipulation of the virtual camera "within" the 3D environmental model, the apparent vantage point of the user may move into and through that depictive space. This allowance is one of the key aspects of "interactivity" in such games (Tavinor 2018, 156).

And yet, most 3D worlds are still compromised by the 2D surface of the depiction; an image on a TV or PC screen, or on a mobile device. The second naturalistic improvement crucial here is the use of binocularity in VR media, which we need to revisit. The binocularity of stereoscopic headsets uses two overlaid perspectives—one for each eye—to give a naturalistic impression of depth to the depictive spaces produced by 3D graphics. The two images are displaced so that by the convergence of the focus of the eyes, the images fuse together, and the visual system interprets the resulting "object" as having a spatial position relative to the apparent position of the user. Hence stereoscopic headsets exploit the visual system to give a naturalistic impression of spatial depth that situates the user within the depictive space. Of course, there were earlier cases of depictive binocularity—famously in the form of *Viewmaster* stereoscopic viewers and 3D films—but these involve the viewer gazing on static scenes or gazing from an apparent static position. The real naturalistic improvement of VR media, therefore, is the combination of the expansion of depictive space via 3D graphics, and the use of binocularity to give these spaces an apparent depth. To these developments owe the overwhelming sense of situatedness or "presence" with which VR media have long been associated (Minsky 1980), and which surely count as their most vivid phenomenological feature.

Even though VR may improve on the limitations inherent in traditional linear perspective, there remain complications and problems. I'll deal with three related problems here: the persisting limited field of view; the lack of focal variation in virtual images; and the so-called accommodation/convergence problem currently faced by stereoscopic headsets. The potential solution to all these problems—the dual rendering of foveal and peripheral vision—will give us a sense of how future VR media could make further naturalistic advances.

The previously noted spatial distortions produced by linear perspective (the unnatural virtual spaces produced at the periphery) are often not a problem in VR because the field of vision is usually limited in a way that they are not apparent (for example PlayStation VR represents a visual field of only 100 degrees). Being effectively "off screen," the distortions are simply not rendered. However, this leads to an important problem with stereoscopic headsets: the field of vision is very obviously constrained, and a user can typically see the border of the frame of the image. Wider fields of vision could reduce the obviousness of the border, but with the potential costs of reintroducing spatial distortions on the periphery of the field of view. Ideally, to replicate the visual field entirely naturalistically, a VR headset would reproduce both foveal and peripheral vision in one array, essentially disguising the "border" of vision much as it is not evident in native vision, a technique called "foveated rendering" (Patney, et al. 2016). This technique necessitates the tracking of eye movement to appropriately render the location of the fixation of foveal vision (Guenter, et al. 2012). This is one key reason why eye tracking is seen as a desirable outcome in current research. (An incidental benefit of this development would be that parts of the scene outside of foveal vision could be rendered in a less detailed way, allowing for savings in processing power.)

The related second problem is that largely because of technical issues, typically in VR headsets everything on the array is rendered as if was the subject of fixated foveal vision. Careful inspection of the image produced by PlayStation VR shows that the entire visual scene—the background, middle ground, foreground, and central and peripheral zones—is crisply focused. Native vision does not present a uniformly distinct and focused visual field, first, because of the already noted distinction between foveal and peripheral vision, and second, as the result of focus and depth of field. The failure to acknowledge these two aspects of native vision gives the visual scene in VR a very unnatural appearance. Again, the effective replication of foveal vision would help here. This would allow only parts of the visual field representing fixated foveal vision to appear to be focused and distinct; correspondingly, both the borders of the visual scene, and those objects at differing focal depths could appear unfocused and indistinct. But the additional complication here is that in addition to eye tracking to detect the fixation of the eyes, the VR system would need to track the *convergence* of the two eyes on specific objects in the visual scene to appropriately place the focal length, and secondly, be able to produce the image appropriate to that specific focal length (that is, with the correct impression of depth of focus).

This leads to the final issue. VR media do not account for accommodation/vergence coupling, that is, the tendency of the eyes to change focal length *and* converge on objects to bring them into focused attention (Hoffman, et al. 2008). Because the actual perceptual object in VR (the screen) is always the same fixed distance from the eyes, and because the stereo displacement of the two images asks the visual convergence system to change when the eyes focus on differently located objects in the virtual space, the two visual systems can be forced to decouple. This decoupling can cause eyestrain and headaches. This is a particularly serious problem with augmented reality, where the focal qualities of virtual objects must be made to *match* the focal qualities of the environment in which they are placed. Again, some form of eye tracking, combined with the ability of the image to render multiple focal planes in one visual scene to allow for depth of field, would be needed to give VR a more naturalistic impression. Nevertheless, should this be achieved it would be a significant contribution to the naturalism of VR media.

It is worth noting that there has been scepticism about the *naturalness* of the techniques of depictive perspective developed by artists, for example from Rudolph Arnheim (1974) and Nelson Goodman (1976). Goodman argued that "the behaviour of light sanctions neither our usual nor any other way of rendering space" (1976, 19). It would seem natural to extend Goodman's claim to the VR depiction of space, and so his critique, if successful, might seem a challenge to the analysis provided here that there is a natural similarity or isomorphism between VR and native visual perspective. I do not have the space to do so here, and the topic is clearly one in need of further work, but it can at least be noted that cognitive science has provided evidence against some of the stronger conventionalist claims of Goodman and his ilk (Kubovy 1986). That the depiction of space depends not only on geometry, but also spatial cues inherent in the psychology of vision, counts against the idea that the depiction of space is entirely *conventional*, because some aspects of a successful spatial depiction will rely on the in-built features of our visual systems. And indeed, I think that consideration of some of the noted challenges to VR naturalism, and the potential solutions to these, further show how VR depictions, in their improvement over previous non-VR modes of spatial depiction, are converging on native perceptual processes much in the way that previous modes of pictorial representation may have been incrementally refined by the process of "schemata and correction" (Gombrich 1960).

5. Virtual Self-Involvement

We have found that one particularly crucial element of the depictions of VR media is the "apparent vantage point" or "point of view," and having this serve as one's apparent situation *within* the experiential space depicted. This is the most vivid and engaging aspect of VR media, but again it is not entirely new, as other artworks, pre-computer, have sometimes depicted the user or viewer as sharing a perceptual space with the depicted environment. A famous case which has already been mentioned here is Jan van Eyck's *Arnolfini Portrait* (1434) where a mirror depicted in the rear of the scene shows the reflection of two viewers—one of whom may be Jan van Eyck himself—looking on the scene depicted in the painting. Jan van Eyck did not employ linear perspective in the *Arnolfini Portrait* because the painting involves several vanishing points. Instead he employed a rather more intuitive sense of perspective which renders a geometrically problematic or chaotic space; however, the impression of a viewer just beyond the pictorial space remains, partly because the depicted figures present themselves to this viewer.

Another example is Diego Velázquez' *Las Meninas* (1656) where the apparent viewers of the scene may be identified with King Philip IV of Spain and his wife who presumably stand just outside of the picture space where the gazes of several of the depicted individuals converge. A further intriguing complication of this painting is that the artist Velasquez is depicted as looking to this couple, and presumably sketching or painting them on a large canvas before him that may itself—though this is disputed—be reflected in a mirror on the far wall. Both paintings thus include an apparent perspective beyond the pictorial space that represents the viewpoint of a diegetic viewer of the scene, and so incorporate the apparent orientation on the picture space of the viewer themselves.

This gives us the background to understand part of the artistic achievements of the *Arnolfini Portrait* and *Las Meninas*. Both paintings extrapolate from formal developments of perspective—in the case of *Las Meninas*, strictly linear perspective and in the *Arnolfini Portrait* a more chaotic sense of space—to realize an artistic technique. In both cases, the depicted space of the painting implies the presence of a viewer within the picture space, and moreover, implies that this viewer occupies the space of the actual viewer. Both works thus connect the viewer with the apparent scene viewed, relating the viewer with the space "in front" of them so that they can become a part of the implied content of the work. Moreover, both works employ this arrangement to develop subtle meanings. In the *Arnolfini Portrait*, the depiction of space and the inference that a viewer of that space exists beyond the picture plane, is exploited to considerable artistic effect. The meaning of the *Arnolfini Portrait* is much contested, but it has been famously (though just a little controversially) suggested by Erwin Panofsky that the viewer—whose presence is depicted in the mirror, but also implied by the pictorial space of the painting—has the role of a witness to a marriage which is depicted in the scene (Panofsky 1934). The meaning of the painting, then, can only be grasped by understanding the spatial implications of the work.

Like the *Arnolfini Portrait* and *Las Meninas*, VR works usually locate their appreciators within the space of their depicted scenes, though not by a trick of mirrors or by spatial inference, rather by freeing the previously fixed perspective, placing it under the control of the appreciator, and associating it with an identity in the diegetic world. The VR medium achieves this not only visually, but also through its use of stereo sound. Through the binocularity of the image and the depicted soundscape, the tracking of the orientation of the user's spatial orientation, and the ability of the user to manipulate the apparent perceptual vantage point of the depicted space, the appreciator finds themselves *within* the apparent spatiality of the depiction, turning their head this way and that, to view and hear the world *around* them, and hence, the apparent relationship that the appreciator has with the depicted scene is deepened. This, as we noted when discussing the Disney research, accounts for the isomorphism in virtue of which stereoscopic headsets count as a virtual medium: the spatiality of the depiction replicates the spatiality of our native perceptual access to the world.

When the perceptual medium is employed as an artistic medium, this *self-involvement* leads to one of the key artistic developments of VR: the user depicted as an agent, individual, or character within the fictional world depicted by the VR media. In *Resident*

Evil VII the player is identified with Ethan Winters, and is given a back story, personal qualities, and various abilities to act in the world of the game. Ethan has been drawn to the deserted house in Louisiana to search for his missing wife. The fiction of *Resident Evil*, while it can be described by observers in the third-person as involving Ethan's subsequent exploration of the house, is also naturally described in the first-person as one's own exploration of the house. Through Ethan, the player makes many things fictionally true of themselves and of their involvement in the game world, for example, that they are "terrified of being alone in the deserted house." Jon Robson and Aaron Meskin argue that videogames are of a kind of "self-involving" interactive fiction where such language is best interpreted as relating what is fictionally true of the player within the game world (Robson and Meskin 2016). Thus, as fictions, videogames warrant that things are fictionally true of the player, and in VR games one means by which they do so is in the depiction of an apparent perceptual orientation by the images in the stereoscopic headset and the stereo soundscape received through the headphones. Adopting Robson and Meskin's terms, VR media allow for a vivid sense of fictive perceptual self-involvement.

The consequent fictional vulnerability of the player-character to threatening events in the game world, and the effect this has on a player's self-concerned emotions, surely comprises an important and unique source of the terror in *Resident Evil VII*. In a survey of the literature on player experience in VR, Dooley Murphy found that one of the most prevalent aspects was what he calls a sense of "patiency," that is, a feeling of having limited agency in the VR world, and a "co-occurrent [. . .] sense of self vulnerability" (Murphy 2017, 10). The sense of perceptual self-involvement provided by VR media is ideal for conveying survival horror fictions such as *Resident Evil* precisely because of how it leaves the player helpless within a threatening world. Furthermore, the expansion of the virtual space to include the *presently unseen environment*, and how this requires an active and embodied perception of the game world, alters the player's epistemic relationship to the depicted world. Rather than a passive relationship where one adopts the gaze inherent in the painter's depiction of the scene, the player must direct their own attention to explore the space. The user's realization that there is a potential depicted space "behind them" and the resulting ability for users to "look behind" them in these virtual spaces lends the self-involved emotions typical of videogames a special immediacy. "What was that sound? Are there monsters back there?! I really hope there aren't monsters back there!" (And, of course, it usually turns out that there are monsters back there.)

Finally, this perceptual and emotional engagement with the fictional worlds of such games stands as one of the best illustrations both of the functional isomorphism in virtue of which the media involved count as virtual media, and the sense in which these media provide a genuinely naturalistic engagement in these fictional scenarios. The perceptual isomorphism generates the possibility of self-directed emotions such as fear, in a way that seems only partially attributable to other media forms such as cinema. In the latter, while one may observe the green slime (Walton 1978) slithering about on the screen, and even fear it slithering towards them, the appreciator is safe from the slime following them as they flee in terror, because such fictions do not allow for the kind of spatial self-involvement allowed in VR fictions. Moreover, from personal experience I judge that such traditional fictions do not give the sense of immediate presence that VR media does.

The development of technology always leads to the question of the extent to which the world after the technology, has been transformed. The claims made about the potential of VR to transform the world, and our understanding of it, have often been quite extreme. Michael Heim considered the medium a "metaphysical machine par excellence" that would usher in a metaphysical revolution requiring us "to dig again in a very ancient form of metaphysics excavated by the engines of computer simulated virtual reality" (Heim 1990, 29). Taking the bait, at least one philosopher has raised the possibility that we already live in a virtual world (Bostrom 2003). Such speculations seem to add little to our understanding of VR, and it is used mostly as a prop to explore familiar Cartesian speculations. In this chapter I have forwarded the deflationary view that VR media are precedented in the pre-VR world, and that rather than an unprecedented or metaphysically provocative revolution, VR media comprise an evolution of existing representational forms such as paintings in linear perspective, 3D movies, videogames, and stereoscopic picture viewers. In this deflationary sense, virtual media employ depictive technologies to give a user an apparent spatial relationship to a depicted space. Indeed, in this sense, Alberti's first depictive spaces really were "virtual spaces"—though limited, imprecise and distorted ones—by depicting those spaces as if the viewer held a specific perspective on the scene. This conclusion strikes me as exactly right: perspectival depiction is a step in the direction of the virtual media we see today. A significant difference is the precise technology used: in Alberti's case, the technology is the geometrical "plan and elevation" method that allows for a 3D spatial projection onto the 2D plane of a painted surface; in VR, it is a computer-dependent technology involving stereoscopic headsets and motion-tracking equipment that allows for a kind of egocentric picturing where one finds their apparent perspective *within* the 3D projection.

Treating VR in this deflationary way—that is, seeing it as an extension of previous representational media—has a variety of implications. I can only gesture toward some of these in this chapter, but they might naturally be the focus of future research. First, the development is interesting because it may initiate artistic devices or experiences that are unseen, or at least rare, in previous media. One example of this is the sense of vulnerability or "patiency" discussed earlier. Another is the barely explored potential of "gestural control" to reconfigure how appreciators of VR artworks might interact with those works. Secondly, the development of VR media may allow us to see how theories and debates focusing on earlier representational media are compromised by their limited focus. For example, in his criticism of the putative spatial realism of linear perspective, Nelson Goodman notes that for linear perspective to achieve any semblance of illusionistic realism in practice, "the picture must be viewed through a peephole, face on, from a certain distance, with one eye closed and the other motionless" (Goodman 1976, 12). But that VR allows for binocularity, the scanning of visual scenes, and the apparent movement of the viewer with respect to the visual scene, completely obviates this criticism; the things that Goodman thinks are missing in the viewing of paintings, are a crucial part of the development of VR media, and VR may be considered a more realistic depictive medium for this very reason (Tavinor 2019b). Similarly, while it might seem a reasonable judgment of earlier pictorial media that pictorial content is "not represented in our egocentric space: the depicted space is not our egocentric space" (Nanay 2015, 189), this is a much less certain claim if extended to the medium of VR, where the depiction of egocentric space is precisely the contribution that VR make to depictive media. The consideration of VR may thus ask us to revisit and reconsider what a representational medium can achieve.

Notes

1. There is also the question of the relationship of virtual media to *computational* or *digital* media, an issue I will not directly tackle here other than to say that I take it that my analysis implies that virtual media need not be computational.
2. The technology can be seen here: https://www.youtube.com/watch?v=Qxu_y8ABajQ
3. This kind of "twofold" seeing is a frequent component of theories of pictorial seeing (e.g., Wollheim 1980), but whether virtual media count as a form of picturing is not a topic that has been given much attention.

References

Arnheim, R. 1974. *Art and Visual Perception: A Psychology of the Creative Eye*. Los Angeles: University of California Press.

Bostrom, N. 2003. "Are You Living in a Computer Simulation?" *Philosophical Quarterly* 53, no. 211: 243–255.

Alistair Burleigh, Robert Pepperell, and Nicole Ruta. 2018. "Natural Perspective: Mapping Visual Space with Art and Science." *Vision* 2 (2).

Chalmers, D. 2003. "The Matrix as Metaphysics." In *Philosophers Explore the Matrix*, edited by C. Grau, 132–176. Oxford, UK: Oxford University Press.

Chalmers, D. 2017. "The Virtual and the Real." *Disputatio* 9 (46): 309–352.

Currie, G. 1999. "Visible Traces: Documentary and the Contents of Photographs." *Journal of Aesthetics and Art Criticism* 57: 285–297.

Edgerton, S. Y. 2009. *The Mirror, the Window & the Telescope: How Renaissance Linear Perspective Changed Our Vision of the Universe*. Ithaca, NY: Cornell University Press.

Gombrich. E. H. 1960. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. Princeton: Princeton University Press.

Goodman, N. 1976. *Languages of Art: An Approach to a Theory of Symbols*. 2nd ed. Indianapolis, IN: Hackett.

Guenter, B., Finch, M., Drucker, S., Tan, D., and Snyder, J. 2012. "Foveated 3D Graphics." *ACM Transactions on Graphics* 31, no. 6: article 164:1–10.

Heim, M. 1990. "The Metaphysics of Virtual Reality." In *Virtual Reality, Theory, Practice and Promise*, edited by S. K. Helsel and J. P. Roth. London: Meckler.

Heim, M. 1993. *The Metaphysics of Virtual Reality*. Oxford: Oxford University Press.

Hoffman, David M., Ahna R. Girshick, and Kurt Akeley. 2008. "Vergence-accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue." *Journal of Vision* 8, no, 3: 33, 1–30.

Hyman, John. 2006. *The Objective Eye: Color, Form, and Reality in the Theory of Art*. Chicago: University of Chicago Press.

Kerlow, I. V. 2000. *The Art of 3D Computer Animation and Imaging*, 2nd ed. New York: Wiley.

Kubovy, M. 1986. *The Psychology of Perspective and Renaissance Art*. New York: Cambridge University Press.

Minsky, M. 1980. "Telepresence," *Omni Magazine* 2: 45–51.

Murphy, D. 2017. "Virtual Reality Is 'Finally Here': A Qualitative Exploration of Formal Determinants of Player Experience in VR." *Proceedings of DiGRA 2017*, Melbourne, Australia.

Nanay, B. 2015. "*Trompe l'oeil* and the Dorsal/Ventral Account of Picture Perception." *Review of Philosophical Psychology* 6: 181–197.

Pan, M., and Niemeyer, G. 2017. "Catching a Real Ball in Virtual Reality." *IEEE Virtual Reality Conference*, Los Angeles, CA.

Panofsky, E. 1934. "Jan van Eyck's Arnolfini Portrait." *The Burlington Magazine for Connoisseurs* 64, no. 372: 117–119.

Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., and Lefohn, A., 2016. "Towards Foveated Rendering for Gaze-Tracked Virtual Reality." *ACM Transactions on Graphics (SIGGRAPH Asia)* 35, no. 6: 1–12.

Peirce, C. S. 1935. *The Collected Works of Charles Sanders Peirce; Volumes IV and V*, edited by Charles Hartshorne and Paul Weiss. Cambridge, MA: Belknap Press.

Robson, J., and Meskin. A. 2016. "Video Games as Self-Involving Fictions." *Journal of Aesthetics and Art Criticism* 74: 165–177.

Scruton, R. 1981. "Photography and Representation." *Critical Inquiry* 7, no. 3: 577–603.

Tavinor, G. 2018. "Videogames and Virtual Media." In "The Aesthetics of Videogames," edited by Jon Robson and Grant Tavinor. New York: Routledge.

Tavinor, G. 2019a. "On Virtual Transparency." *Journal of Aesthetics and Art Criticism* 77, no. 2: 145–156.

Tavinor, G. 2019b. "Towards an Analysis of Virtual Realism." *Proceedings of DiGRA 2019*, Kyoto, Japan.

Walton, K. 1978. "Fearing Fictions." *Journal of Philosophy* 75, no. 1: 5–27.

Walton, K. 1984. "Transparent Pictures: On the Nature of Photographic Realism." *Critical Inquiry* 11, no. 2: 246–277.

Walton, K. 1990. *Mimesis as Make Believe*. Cambridge, MA: Cambridge University Press.

Wollheim, R. 1980. "Seeing-as, Seeing-in, and Pictorial Representation." In *Art and its Objects*, 2nd ed., edited by R. Wollheim, 205–226. Cambridge: Cambridge University Press.

CHAPTER 22

EVALUATION, VALIDATION, AND MANAGEMENT IN DESIGN

PIETER E. VERMAAS

1. Introduction

The question of whether technologies realize the aims for which they are developed is one that in philosophy of technology is regularly answered negatively. The overall development of technology has been analyzed by some as having an internal deterministic logic, thus denying that aims of people, firms, or societies can steer this development (e.g., Ellul 1962). Individual technological projects defined by specific aims have been disclosed as misconceived 'technological fixes' by revealing that the aims were eventually not realized (e.g., Volti 1992; Rosner 2004). And social processes between stakeholders may guide the aims that are in the end attached to technologies, and thus overrule the intentions of their designers (e.g., Pinch and Bijker 1987). These answers suggest that technologies come out as poor means when evaluated against the aims for which they are developed. And these answers suggest that we—philosophers of technology, and others—had better closely and critically monitor newly developing technologies, as we currently do with, for example, gene editing and autonomous vehicle technologies.

The question of whether engineering design realizes the tasks it takes up seems, in contrast, to have a positive answer. These tasks originate from clients—individual persons, firms, or governmental agencies—and engineering design has developed effective practices, tools, and methods to create technical artifacts, or more precisely, descriptions of technical artifacts, that realize these tasks. Clearly some engineering design projects fail, and some tasks may turn out to be unrealizable—think of the challenge to create nuclear fusion energy plants. But on average engineering design comes out as effective: it often realizes the tasks it takes up, and has thus built up an extensive track record of successes for its practices, tools, and methods. We can, of course, still scrutinize engineering design closely and critically, yet this seems not to be a first priority when monitoring whether technologies realize our aims.

This discrepancy in the evaluation of technologies and of engineering design has an explanation. In engineering design the task taken up is to describe technical artifacts that meet specific well-defined design requirements. These requirements capture the aims the clients have but typically in a rather restricted way: they fix the artifacts the clients want in terms of their structural and functional properties, and the means and costs it takes to manufacture them. These design requirements need not fix, however, the impact of the designed artifacts when offered to users and society, for example, whether the artifacts are successful in the market place, how users will employ the artifacts in the long run, and what intended or unintended impact the use of the artifacts have on society. By focusing on (only) the more restrictive design requirements, engineering design projects can be managed and concluded in well-defined ways: during a project the requirements provide a basis to determine if progress is made, and whether the final design concludes the project successfully. And with its focus on the design requirements, engineering design can show again and again that it realizes the aims of its clients. The evaluation of the designed artifacts as technologies that are made available to users and society involves in contrast their broader impact on users and society. This impact need not meet the expectations of clients, even if the technical artifacts do satisfy the design requirements.

This explanation is, I agree, a caricature. Yet I hold it as relevant to the current evolution of engineering design into *design thinking* (e.g., Martin 2009; Lewrick et al. 2018). It is a caricature because engineering designers already look beyond their restrictions by helping clients articulate and detail their aims and by anticipating and avoiding misuses of the designed artifacts. It is relevant, though, because with design thinking designers are more substantially abandoning the restrictions of engineering design. Design thinking includes, for instance, innovative design where, adopting the position ascribed to Henry Ford that "[i]f I had asked people what they wanted, they would have said faster horses," it is assumed that true innovation is arrived at when designers take over from clients the role of defining the aims designed for. And innovative design disrupts users by not offering artifacts with predictable uses but by creating 'game changers' that overhaul current uses. In innovative design, and design thinking in general, a 'thinking out of the box' approach is replacing the restrictive engineering design approach. On this view, designers should actively analyze the aims and situations of their clients and of the users, look for possibilities to innovate by reformulating the aims and specifications clients initially come up with, and create novel and surprising designs and uses.

The position I take in this chapter is that with this innovative, thinking-out-of-the-box approach, design thinking confronts us with similar evaluation problems as technologies do. By this approach designers create design projects in a more independent fashion, one which we currently lack the means to manage and evaluate. This suggests that design thinking should be monitored as closely and critically as technologies are. My position is however not one of rejection. Design thinking does lead to innovation (e.g., Brown 2009; Verganti 2009) and may be valued for that. And design thinking is leading to other interesting forms of design. It includes, for instance, *social design* where designers take up societal challenges and aim at designs that resolve them through behavioral changes in people (e.g., Marzano 2007; Brown and Wyatt 2010). An example is 'nudging,' which involves designing choice architectures for steering the decisions of people in specific directions (Thaler and Sunstein 2008). Design thinking also includes *value-sensitive design* (Friedman et al. 2006) and *design for values* (Van den Hoven et al. 2015), in which designers accept moral and societal values of stakeholders as additional design requirements. These are in my view all valuable new forms of design that should be developed further. Yet for all these forms of design thinking we currently neither have clear means for evaluating whether they lead to successful outcomes, nor long track records by which we can conclude that their practices, tools, and methods guarantee such successes. I moreover will argue that the evaluation and validation of design thinking are topics philosophy of technology should be concerned with. Participating in the development of this evaluation and validation involves analyzing and understanding how design thinking is structured, and will turn design thinking into a reliable means for realizing the normative aims that philosophers of technology are also committed to. Still, as long as we do not have these means for evaluating efficacy, we had indeed better monitor design thinking closely and critically.

In this contribution I review the evaluation and management of engineering design and design thinking, respectively. section 2 compares the evaluation of the designs created by engineering design and design thinking, while Section 3 considers the validation of the design methods applied in engineering design and design thinking. When considering engineering design I focus on the design of material artifacts rather than software. When considering design thinking I focus on the design of innovative solutions rather than of incremental improvements of existing solutions. Finally, I take effectiveness as the central criterion of evaluation, yet in the concluding section 4, I broaden the analysis briefly to the criterion of efficiency. In that section I also argue that philosophy of technology should be involved in the evaluation and validation of design thinking.

2. Evaluation of Designs

A design of a technical artifact can be evaluated by numerous criteria. An obvious one is that it meets the design task that is set by the designer for capturing the aim of the client. Yet as will become clear in this section, this criterion is not so straightforward in the case of the innovative 'thinking-out-of-the-box' approach of design thinking. With that approach, designers may reformulate the clients' initial wishes, if there is a client at all, meaning that the designed artifact should just meet the task as set by the designer. This design task criterion can in both cases be articulated in material terms (e.g., dimensions and materials), in functional terms (what it should do), and in teleological terms (what can be achieved with it in use). Other criteria for a design are that it is optimized, economic, useable and ergonomic, maintainable, producible, recyclable, safe, sustainable, and compliant with relevant regulations and laws.

In engineering design it is indeed the client who determines the design task, and the engineering designer who has to deliver a design that meets that task. In the methodological account of engineering design by Pahl and Beitz (Pahl et al. 2007), the client's aims as well as many of the other criteria mentioned above are brought together in one overall *design requirement list*. This list can be divided into different sublists grouping together requirements of the same type (say, functional requirements or safety requirements) and requirements may be differentiated (say, as demands or as wishes). This list acts as an overall criterion that the final design should meet to be evaluated as successful. Using this account of engineering design, I focus in this section only on the evaluation of a design with the criterion of effectiveness, by which I mean that the final designed artifact should meet all requirements in the requirement list. This means that I am ignoring in this contribution the problem of having to make trade-offs in the evaluation of designs, as when a design meets some requirements well and others less well.

On the account by Pahl and Beitz, engineering design projects are divided into four steps: *task clarification*, *conceptual design*, *embodiment design*, and *detail design*. The requirement list plays a central role in these steps. In the task clarification step, designers identify the requirements the design should meet in order to realize the client's aims. This identification leads to the mentioned requirement list, "against which the success of the design project can be judged" (Pahl et al. 2007, 145). The list is however not fixed within the task clarification step but can be amended and extended later in the project, when new information about the design solution becomes available. It is described as a "binding yet provisional" requirement list (Pahl et al. 2007, 155). In the conceptual design step principle solutions are determined for the design. This is done by exploring and then choosing a function structure of the artifact-to-be, fixing working principles for delivering the functions, and composing these working principles into a working structure for the design. The requirement list serves as inspirational input to conceptual design and acts as a criterion the found principle solutions should minimally meet to be good (Pahl et al. 2007, 192–194). In the third, embodiment design step, the principle solutions are developed into an overall spatial layout design of the technical artifact, including designs of the component shapes and materials. Again the requirement list is a source for this design by containing demands and wishes for spatial sizes and materials to be used. As before, the resulting embodiment designs should minimally meet the requirement list (Pahl et al. 2007, 416–417). The final detail design step completes the embodiment design step, with final specifications of the components in terms of their shapes, forms, et cetera, and instructions about how to manufacture them.

The requirement list gives on this account a clear means for evaluating the resulting designs at each step, and also for evaluating and managing the process of engineering design projects. This list is drawn up at the beginning of the design process for capturing the aims of the client. So the client can both check initially if this list is indeed corresponding to his or her aims, and determine at the end if the design is effective as an artifact that meets the requirements on this list. The list may, as said, be amended and extended and thus change during the design process, yet these changes are made explicit such that the client may in principle review them. The requirement list also gives the designers, client, and other stakeholders means to manage the design process, although in a coarse-grained manner. The conceptual and embodiment design steps in engineering design projects are concluded if their results—the principled solutions and the spatial/material design, respectively—meet the listed requirements. Hence with the requirement list one also can determine that steps in design projects are concluded successfully.

In design thinking, equivalents to the requirement list of engineering design do not exist. In the extreme case a design thinking project does not even have a client, as when designers themselves pick up a challenge for which they see opportunities to come up with an innovative solution. When a design thinking project does have its starting point in an aim of a client, the thinking-out-of-the-box approach stimulates designers to not take that aim at face value, but to first do research on the client and aim to find the 'real underlying problem,' reframe the problem in a more productive form, or otherwise change that aim. There are multiple design methods proposed for design thinking, and they are also structured by steps, as methods for engineering design are. But design thinking methods start with steps such as *empathize* and *define* (D.school 2018), *understand*, *observe*, and *point of view*, (Plattner et al. 2009) and *archeology*, *context*, and *frame* (Dorst 2015). By these steps designers are stimulated to get some distance from the aims the client comes up with, and to understand the aim, explore the context, and eventually frame it using a personal point of view.

The power of this thinking-out-of-the-box approach, as well as the problem it creates for the evaluation of such design projects, can be illustrated by the example of the redesign of the Kings Cross entertainment district in Sydney, Australia (Dorst 2011, 528–530; Dorst 2013; as earlier analyzed in Vermaas et al. 2015). The Kings Cross district posed a law-and-order problem the City of Sydney had difficulty managing:

> Being the main night-time entertainment district in Sydney, Kings Cross has increasingly become a setting for antisocial behaviors and escalating crime. High volumes of young people attend on Friday and Saturday nights, and activities are predominantly concentrated on a small stretch of nightclubs. Some of the problems that occurred include drunkenness, violence, petty theft, and drug dealing. Previous attempts at solving the problem by the City of Sydney included the implementation of strong-arm tactics and the increasing of police presence; however, the additional security measures failed to enhance feelings of public safety and instead resulted in a grim atmosphere for all.
>
> (Vermaas et al. 2015, 134)

With the recognition that further policing measures were not addressing the problem, the City of Sydney turned in 2008 to the Designing Out Crime Research Centre. http://designingoutcrime.com/">Designers in this center were asked to explore the situation in the Kings Cross district, and define opportunities for reducing crimes and misdemeanors, in particular 'alcohol-related violence.' Using the tools of design thinking these designers arrived at an alternative analysis of the problem that the Kings Cross district posed:

> The designers concerned quickly realized that the situation had previously been treated as a law-and-order problem requiring law-and-order solutions; however, the people involved were not actually criminals. Instead, they were just young people looking to position themselves in a social setting and to have a good time. The lack of structure of the nightspot together with the sheer volume of young people meant that they were becoming bored and frustrated, and consequently were not having a good experience at all—a problem only exacerbated by the additional security measures. The designers proposed a simple analogy in which large volumes of people already successfully come together and interact in a harmonious fashion: a music festival. They effectively reframed the problem by comparing the dysfunctional situation at Kings Cross with a well-organized music festival. They asked themselves what they would do if they were organizing a music festival and this triggered new scenarios for action, as a well-organized music festival offers many facilities that are not currently available in the Kings Cross district but could easily be designed in.
>
> (Vermaas et al. 2015, 134–135)

This reframing of the problem allowed the designers to explore a series of new solution directions, and propose measures to improve the situation at Kings Cross. The reframing allowed, for instance, to look at the way in which visitors travel to and from the Kings Cross district. In a regular music festival, people can arrive and leave when they want. At Kings Cross this opportunity was discovered to be less available. Services to the train station at Kings Cross ended around the time at night that visitors travelled to the district, making these services also useless to visitors who wanted to leave. This discovery enabled the designers to propose measures to make later public transport available to visitors. The observation that the existing security measures at Kings Cross led to a grim atmosphere, allowed exploring a second solution direction of adding friendly 'Kings Cross Guides' to the existing presence of police officers and bouncers. These guides could welcome visitors into the area, help visitors by providing information on all the facilities, and also warn the police before a situation may get out of hand. In all, the 'music festival' frame allowed the designers to explore about twenty solution directions, of which many have been tested and implemented.

This example illustrates the power of the thinking-out-of-the-box approach of design thinking. By reframing the aim of the client, designers can find new solutions to respond to this aim. The problem it creates for evaluating the new solutions is that it is not immediately clear that these new solutions are actually realizing the original aim that the client had in mind. In short: the City of Sydney wanted to reduce crimes and misdemeanors in the Kings Cross area; the designers improved the quality of the entertainment service in this area; and it is not evident that the second realizes the first. It may, of course, be plausible that if young people are welcomed by friendly guides, can have a good time at Kings Cross, and leave smoothly if they want to, these young people will then be less inclined to engage in "drunkenness, violence, petty theft, and drug dealing." Yet it might also not be so.

There are at least two options for evaluating individual designs that come out of design thinking projects, such as the one in Kings Cross. The first option stays close to engineering design and puts emphasis on the original aim of the client: just as the requirement lists capture that aim in engineering design, one should find criteria for evaluating the final designs of design thinking projects against the original client's aim. At the end of this section I will sketch what this first option could amount to. Yet in the literature on design thinking, this option seems less accepted since it imposes constraints upon the innovation that design thinking can bring. The client is presented as conservative and less informed, as captured by the Henry Ford quote. And innovation is presented as a non-linear process, one that does not flourish with early external management or imposed business models (Kyffin and Gardien 2009, 57–58). The second option avoids such constraints and focuses on the formulation of the design task as eventually developed by the *designers*: the task as adopted by the designers defines what the project is about, hence a design thinking project should be evaluated against *that* task. Research on how individual designs of this type can be evaluated has led to two initial proposals, for example, in social design and in "serious games" design for health care.

Evaluation of the designs that are produced by social design methods is confronted with the problem that such design typically aims at behavioral changes of people within specific (social) contexts, changes that may surface only after extended periods. This problem emerges in the Kings Cross example, and in related design thinking fields such as nudging. For determining whether a social design may have its desired effect in the long run, short term laboratory and field studies are not meaningful. Yet what can be achieved is a more qualitative assessment by developing a narrative about how the design is expected to realize its effects, and letting this narrative be assessed for causal coherence and plausibility by experts active in the relevant social domains (Tromp and Hekkert 2016).

For serious game design for health care purposes—think of games for abstaining from unhealthy behaviors such as smoking, and for adhering to medication use—the problem of evaluating effectiveness over time is observed as well. Yet possibly due to the high thresholds in health care for allowing new therapies, work has been done on making the behavioral effects of (playing) serious games more directly measurable and provable. In (Graafland et al. 2014), for instance, a framework is presented for describing serious games and their assessment, borrowing concepts for measuring behavior from psychology. Van der Kooij et al. (2015), accepting the prevailing use of randomized controlled trials in health care, discuss how this validation method can be applied to demonstrating the effects of serious games. This application leads to methodological problems, for example the choice of control groups. Yet these problems do not entail that it is untenable to require evaluation of serious games. Solutions are available; for example, a "placebo game" can be used for the control group when understanding is sought of what factors in the tested game created the desired behavior change, and the usual medical therapy can be used for the control group when it should be demonstrated that the serious game is presenting an improvement in health care.

This work on evaluation may evolve into a more established and accepted practice of evaluating designs created by design thinking. Still, with this second option, evaluation is performed against the design task as determined by the designers, opening the possibility that designers realize primarily their own tasks rather than the aims of their clients. This result may be defended by pointing to the conservatism of clients or to the benefits of innovation. It however also introduces risks for design thinking projects, as can be illustrated with the example of the Kings Cross redesign project.

Let us return to that project and assume that it successfully realized its self-set task of raising the quality of the entertainment service. Let us also assume that this higher quality reduces crimes and misdemeanors by the visitors of the Kings Cross area. It then follows that some crimes and misdemeanors will remain and are not eliminated by the solutions developed in the redesign project. This became clear in 2012 and 2013 with two separate tragic events in which two young men died by attacks by impulsive aggressors. These events shocked public opinion and resulted in direct governmental intervention to severely limit the opening times of bars and clubs in the Kings Cross area. With this intervention the Australian Government went back to its original frame of seeing the Kings Cross district as primarily a law-and-order problem, meaning that the music-festival frame as proposed by the designers was overruled. This overruling can be interpreted as entailing that in the end, the solutions given by the Kings Cross redesign project should be evaluated as unsuccessful. Improving quality of the entertainment service was in the end not what was needed; the ultimate aim remained reducing crime and misdemeanors. And when it became clear that this aim was not fully realized, the redesign project was rejected. Or, more generally, a client may follow designers in the reframing of his or her original aim, yet this original aim still can play a decisive role in the evaluation of the resulting design solutions by the client.

This negative conclusion about the Kings Cross example should in my opinion not be taken as reason to distrust design thinking; it rather shows that the first option of evaluating the outcomes of design thinking projects against the original aims of clients should be developed as well. In (Vermaas et al. 2015) this option is analyzed by exploring criteria that reframing should meet. Two such criteria are proposed. First, it should be shown that by completing the reframed task the original client aim is also realized. Second, the solutions found through the reframing of the aim should be acceptable to the client. For the Kings Cross redesign project the first criterion implies that it should be shown that an improvement of the quality of the entertainment service indeed reduces the rate of crimes and misdemeanors. The second criterion implies that it should be plausible that the City of Sydney can abandon its traditional role of law enforcer and adopt the new role of host. It is the second criterion that seems not to have been met in the Kings Cross redesign project. Further research should focus on additional criteria to evaluate design thinking projects against the original aims of clients.

3. Validation of Design Methods

In the previous section it was argued that engineering design has in the form of requirement lists the means to evaluate whether the designs it creates realize the aims of clients for whom these designs are created. Furthermore, since engineering design has already been practiced for multiple decades, one can add that it has built up an extensive track record of successfully concluded design projects. Hence, engineering design can present itself as a means to realize client aims. It can present itself as having practices, tools, and design methods, such as the ones given by Pahl and Beitz, that are sufficiently validated over the years, and that are thus guarantees for the effectiveness of future engineering design projects. Design thinking is not in this position. It does not yet have clear checks for evaluating the designs it creates, and since it is a practice that emerged only in the current century, it has not yet had the opportunity to build up a substantial track record of effectiveness. Proponents of design thinking may still claim that their methods of design thinking almost always lead to successful innovation (e.g., Plattner et al. 2009, 103). Yet the basis of such claims is unclear. 'Successful innovation' is, for instance, a rather ambiguous label, making it unclear how to apply it to projects. Did, for example, the Kings Cross redesign project lead to successful innovation? One can argue that it did, because the project created many innovative design solutions for the Kings Cross entertainment district, which moreover found their way to other entertainment areas (Vermaas et al. 2015). At the same time, one can argue that this project was an unsuccessful innovation since its solutions were eventually rejected for the Kings Cross district. One may resolve this ambiguity by noting that innovation is currently often used to refer to the *process* of developing new and promising ideas, by which design thinking may be taken as always innovative. The label can however also be used for singling out projects that actually lead to innovative *outcomes*, so-called 'disruptive game changers' in the world. In that case innovation is a much rarer phenomenon (e.g., Kyffin and Gardien 2009, 57), and design thinking is probably only occasionally innovative.

A way forward to establishing the effectiveness of design thinking is to more directly validate its design methods as guarantees. Design methods in engineering design may get their status of being validated through their track records of successful design projects. Yet an alternative route towards validation is through direct research and testing. In the design research literature, proposals have been formulated for making this route available (Seepersad et al. 2006; Blessing and Chakrabarti 2009). This latter route offers opportunities to validate design thinking's methods more quickly than through the accumulation of extensive track records. In this section I therefore discuss a few of the proposals for the validation of design methods.

The validation of design methods is neither a practice that is established in design research, nor a topic on which research is converging to some sort of consensus. But despite this lack of uptake, it is a topic that finds its way into the design research literature with some regularity. Frey and Dym (2006) give an early overview and explore how

Table 22.1. The validation square

Quadrant 1 Showing internal consistency of the elements that make up design method D	Quadrant 4 Arguing that the effectiveness of D holds for the whole application domain of D
Quadrant 2 Choosing example problems that represent the application domain of D	Quadrant 3 Determining effectiveness of D for solving the example problems

the analogy with validation of medical treatments in health care can provide tools for validating design methods. Although they argue that the analogy does not hold in all respects, they propose for design method validation tools such as clinical randomized controlled trials and natural experiments, in which the effects of the use of design methods are assessed in broad studies of actual practices.

A fairly detailed proposal for validating the effectiveness of design methods is the Validation Square method (Seepersad et al. 2006). This proposal consists of a number of steps that are ordered in four quadrants (see Table 22.1), explaining its name. The first quadrant captures steps that are aimed at establishing the internal consistency of the design method. A design method is in the proposal taken as built up of elements from other methods, and these first steps are about checking the consistency of these elements and of their integration in the method. The second quadrant consists of choosing *example design problems* that represent the domain for which the design method is meant to be effective. By focusing on these example problems, one avoids the need to consider how the design method applies to all problems it is intended for. Yet it should be evident that the example problems are indeed characteristic for the method's domain. Third, it must be established that application of the design method to the example problems indeed resolves these problems, and that this resolution is due to the design method. The fourth quadrant consists of an argued 'leap of faith' that the effectiveness of the design method for addressing the example problems can be accepted as establishing effectiveness for all the design problems in the domain for which the method is meant.

Blessing and Chakrabarti (2009) propose a general research method, called DRM (Design Research Methodology), for research on design tools and methods. This research method contains four stages: (1) a *research clarification* stage aimed at identifying the aspect of design that is to be improved by a design tool by giving a description of the existing situation and of the desired situation; (2) a *descriptive study I* stage in which the literature is reviewed, possibly accompanied by some research, for arriving at an understanding of the factors through which the existing situation could be changed to the desired situation; (3) a *prescriptive study* stage for determining the factor the design tool should support for changing the existing situation to the desired situation; and (4) a *descriptive study II* stage for testing the design tool for its ability to realize the desired situation.

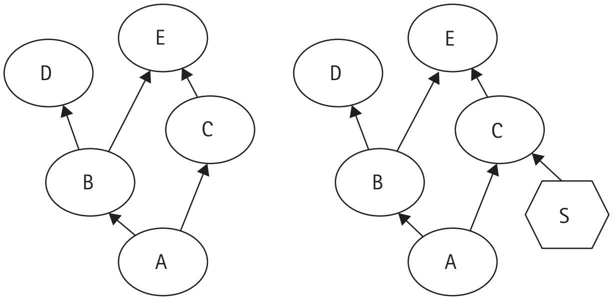

FIGURE 22.1: Examples of a reference model (left) and an impact model (right) in DRM; the hexagonal node in the impact model represents the added support S and the bold arrows and nodes represent the causal chain leading to the improvement of E.

In DRM two models play a central role: a reference model and an impact model. Both models are networks of influencing factors, containing nodes that represent factors as aspects of design that influence other aspects of design, and containing arrows that represent how the factors causally influence each other (see Figure 22.1). The reference model represents the existing situation in design and acts as a benchmark for the improvements. The impact model represents the desired situation and adds the planned support for creating the improvements as an additional factor relative to the reference model.

The fourth stage, descriptive study II, is one in which the efficacy of the design tool is tested by determining how the tool changes the existing situation. It is unclear whether this stage can also be taken as one in which *efficiency* of the tool is tested. If the existing situation in design practices, as captured by the reference model, is interpreted as a situation in which existing design methods are used, one could argue that the descriptive study II stage benchmarks the new design tool against these existing design methods. I return to this point in the concluding section.

A final possibility for validating design methods can be formulated by returning to the methodological account of engineering design by Pahl and Beitz as described in section 2. On that account, engineering design projects are divided into four methodological steps—*task clarification*, *conceptual design*, *embodiment design*, and *detail design*—which all have relatively well-defined inputs and outputs. With these inputs and outputs, of which the requirement list is a key element, it can be evaluated whether these steps are concluded successfully, enabling a transparent tool for managing engineering design projects by the designers themselves, by clients, and by external managers. The different design methods for design thinking also divide design projects into methodological steps, yet now with more qualitatively and vaguely defined inputs and outputs. In (Vermaas 2013) it is proposed to articulate the input and output of these methodological steps in more detail, in order to arrive at a better understanding of how design thinking is structured, to allow transparent management of design thinking projects, and to give clients and other stakeholders the means to evaluate these projects. With this articulation one can also envisage research on the efficacy of design thinking methods by reviewing how often and how fast designers manage in design thinking projects to conclude the methodological steps advanced by the design thinking methods. In (Thurgood et al. 2015) this articulation is given for three design thinking methods.

The validation of the design methods for design thinking requires greater specificity about design thinking, just as is required for the evaluation of the designs created with design thinking. For instance, when randomized controlled trials are adopted for validating design thinking, the "non-design thinking" contrast class has to be determined, and claims have to be formulated regarding in what respect design thinking performs better than non-design thinking. This formulation requires getting specific about how regularly design thinking leads to innovation. Do design thinking methods almost always lead to innovative solutions, as advanced by (Plattner et al. 2009) or do they only lead in some projects to innovation as compared to non-design thinking methods? Likewise, when design thinking methods are validated by the Validation Square method, example problems have to be given that can be taken as representative of the domain of tasks that design thinking is meant to tackle. Instead of seeing it as a general approach to innovation, design thinking is then better presented (and detailed) as a family of separate approaches for different domains, as product development, service design, social design, institutional design, et cetera. Validation with DRM would push design thinking even more towards making explicit what aspects of design it intends to improve, and by what mechanisms it wants to do so. Finally, for validation of design thinking methods through the evaluation of the individual steps of the methods, it will be necessary for the internal structure of design thinking to be clarified in detail. Providing this required clarity probably will only strengthen the case of design thinking, since it would make design thinking more intelligible. Yet this also may be a hurdle, since articulation may be hard to do and may be interpreted by design thinking proponents as a step towards making design thinking as mechanical as engineering design.

4. Discussion and Conclusion

I opened this contribution with the warning that philosophers of technology had better closely and critically monitor design thinking because we currently lack the means for determining whether design thinking realizes the aims it promises to deliver. For engineering design, on the other hand, we have criteria to evaluate the designs it creates and we have longer track records of successful designs by which we can assume that the practices, tools, and methods of engineering design guarantee equal successes for future design projects. For design thinking we neither have such evaluation criteria nor long track records that support that their practices, tools, and methods guarantee successes. In section 2 it was shown that in engineering design, requirement lists can be used to evaluate whether the designs it creates realize the aims of clients. In design research, work is also done to arrive at a means to evaluate the designs created by design thinking. There are two options for this: evaluation of the designs created by design thinking against the aims of clients, and evaluation of these designs against the tasks the designers define. In section 3 proposals for validating design methods through direct research rather than through track records were discussed. Direct validation of the design methods of design thinking may provide alternative guarantees that design thinking will deliver what it promises to deliver. A precondition to this direct validation is that design thinking methods and their promises are articulated with more precision.

These explorations may give reason to think that my warning can be retracted in the near future, when design researchers have reached consensus about the criteria for evaluating the designs of design thinking, and when the design methods of design thinking are validated for their effectiveness. One may then argue that still more work has to be done, since these criteria and validation concern only the effectiveness of design thinking. Another central value in technology is efficiency, and applying this value to design thinking means that we also need the means to determine which design methods are quicker to lead to innovative solutions in product development, make it easier to accomplish behavioral changes in social design and nudging efforts, and lead more rapidly to the incorporation of our values in technologies with the design for values approach. In section 3 it was noted that the DRM approach to validating design methods may be interpreted as also establishing this efficiency of methods. The assessment of the efficiency of design tools and methods is, however, not common in design research. Design tools and methods are typically presented and used side-by-side (e.g., Kumar 2013) without comparing them explicitly or benchmarking them against each other. Initial work on this benchmarking can be found in (Bohm et al. 2017). The comparison of design methods for their efficiency may however be seen as a topic that can be left to design researchers; to retract my warning that philosophers of technology should monitor design thinking, it will suffice that its effectiveness can be established.

One may hold that the evaluation of designs and the validation of design methods are also primarily topics for design research, and not among the ones that philosophy of technology should be concerned with. I believe that there are at least three reasons for philosophy of technology to get involved in this research. The first reason is an epistemic one. Research to enable evaluation and validation needs to make explicit how design thinking is structured and what it can deliver. As such, this research amounts to *analysis and articulation* of design thinking, to which philosophy of technology—more precisely, philosophy of engineering—can contribute and learn from.

A second reason is one of co-responsibility. Design thinking involves social design, nudging, and design for values, which all address normative issues we in philosophy of technology, as well as those in environmental and social philosophy, are also concerned with. With supporting developments such as responsible research and innovation (RRI) and design for values, philosophy of technology is committed to making design thinking an effective means for guiding technological development by moral and societal values.

The third reason is related to the second and can be taken as ethical. Philosophy of technology has criticized unconditional faith in technologies, and revealed how so-called 'technological fixes' led to results we did not aim at. A similar unconditional faith in design thinking should be avoided. Design thinking is currently advanced as a promising means for RRI, social improvement, and design for values. This new faith should be critically examined before it is embraced fully; otherwise design thinking may evolve into the technological fix, or design fix, of the twenty-first century.

References

Blessing, Luciënne T. M., and Amaresh Chakrabarti. 2009. *DRM: A Design Research Methodology*. London: Springer.

Bohm, Matthew, Claudia Eckert, Chiradeep Sen, Venkatamaran Srinivasan, Joshua D. Summers, and Pieter Vermaas. 2017. "Guest Editorial: Thoughts on Benchmarking of Function Modeling: Why and How." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 31: 393–400.

Brown, Tim. 2009. *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*. New York City, NY: Harper Business.

Brown, Tim, and Jocelyn Wyatt. 2010. "Design Thinking for Social Innovation." *Stanford Social Innovation Review* (Winter): 30–35. http://www.ssireview.org/articles/entry/design_thinking_for_social_innovation/

Dorst, Kees. 2011. "The Core of 'Design Thinking' and Its Application." *Design Studies* 32: 521–532.

Dorst, Kees. 2013. "Shaping the Design Research Revolution." In *Proceedings of the 19th International Conference on Engineering Design*, Vol. DS75-01: Design for Harmonies. Seoul, 173–182. Design Society.

Dorst, Kees. 2015. *Frame Innovation: Create New Thinking by Design*. Cambridge, MA: MIT Press.

D. School. 2018. "D.School Bootcamp Bootleg." Hasso Plattner Institute of Design at Stanford. Retrieved 22 July 2018. https://dschool.stanford.edu/resources/the-bootcamp-bootleg

Ellul, Jacques. 1962. "The Technological Order." *Technology and Culture* 3: 394–421.

Frey, Daniel D., and Clive L. Dym. 2006. "Validation of Design Methods: Lessons from Medicine." *Research in Engineering Design* 17: 45–57.

Friedman, Batya, Peter H. Kahn Jr., and Alan Borning. 2006. "Value Sensitive Design and Information Systems." In *Human-Computer Interaction in Management Information Systems: Foundations*, edited by Ping Zhang and Dennis Galletta, 348–372. Armonk, NY: M. E. Sharpe.

Graafland, Maurits, Mary Dankbaar, Agali Mert, Joep Lagro, Laura De Wit-Zuurendonk, Stephanie Schuit, Alma Schaafstal, and Marlies Schijven. 2014. "How to Systematically Assess Serious Games Applied to Health Care." *Journal of Medical Internet Research; Serious Games* 2: e11.

Kumar, Vijay. 2013. *101 Design Methods: A Structured Approach for Driving Innovation in Your Organization*. Hoboken, NJ: Wiley.

Kyffin, Steven, and Paul Gardien. 2009. "Navigating the Innovation Matrix: An Approach to Design-Led Innovation." *International Journal of Design* 3: 57–69.

Lewrick, Micheal, Patrick Link, and Larry Leifer. 2018. *The Design Thinking Playbook: Mindful Digital Transformations of Teams, Products, Services, Businesses and Ecosystems*. Hoboken, NJ: Wiley.

Martin, Roger. 2009. *The Design of Business: Why Design Thinking is the Next Competitive Advantage*. Cambridge, MA: Harvard Business Press.

Marzano, Stefano. 2007. *Flying over Las Vegas*. Eindhoven: Koninklijke Philips Electronics NV.

Pahl, Gerhard, Wolfgang Beitz, Jörg Feldhusen, and Karl-Heinrich Grote. 2007. *Engineering Design: A Systematic Approach*, 3rd ed. London: Springer.

Pinch, Trevor J., and Wiebe E. Bijker. 1987. "The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." In *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, edited by Wiebe E. Bijker, Thomas P. Hughes, and Trevor J. Pinch, 17–50. Cambridge, MA: MIT Press.

Plattner, Hasso, Christoph Meinel, and Ulrich Weinberg. 2009. *Design Thinking: Innovation Lernen—Ideenwelten* Ö*ffnen*. Munich: mi-Wirtschaftsbuch.

Rosner, Lisa, editor. 2004. *The Technological Fix: How People Use Technology to Create and Solve Problems*. New York City, NY: Routledge.

Seepersad, Carolyn C., Kjartan Pedersen, Jan Emblemsvåg, Reid Bailey, Janet K. Allen, and Farrokh Mistree. 2006. "The Validation Square: How Does One Verify and Validate a Design Method?" In *Decision Making in Engineering Design*, edited by Kemper E. Lewis, Wei Chen, and Linda C. Schmidt, 303–314. ASME.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Thurgood, Clementine, Kees Dorst, Sam Bucolo, Mieke van der Bijl-Brouwer, and Pieter Vermaas. 2015. "Design Innovation for Societal and Business Change." In *Proceedings of the 20th International Conference on Engineering Design (ICED15)*, July 27–30, 2015, Milan, Italy, Vol. 8: Innovation and Creativity, 61–70. Design Society.

Tromp, Nynke, and Paul Hekkert. 2016. "Assessing Methods for Effect-Driven Design: Evaluation of a Social Design Method." *Design Studies* 43: 24–47.

Van den Hoven, Jeroen, Pieter E. Vermaas, and Ibo Van de Poel, editors. 2015. *Handbook of Ethics, Values and Technological Design*. Dordrecht: Springer.

Van der Kooij, Katinka, Evert Hoogendoorn, Renske Spijkerman, and Valentijn Visch. 2015. "Validation of Games for Behavioral Change: Connecting the Playful and Serious." *International Journal of Serious Games* 2: 63–75.

Verganti, Roberto. 2009. *Design Driven Innovation: Changing the Rules of Competition by Radically Innovating What Things Mean*. Boston, MA: Harvard Business Press.

Vermaas, Pieter E. 2013. "On Managing Innovative Design Projects Methodologically: The Case of Framing." In *Proceedings of the 2nd Cambridge Academic Design Management Conference*, September 4–5, 2013, Cambridge, UK, 549–560. http://www.cadmc.org/CADMC2013Proceedings.pdf

Vermaas, Pieter, Kees Dorst, and Clementine Thurgood. 2015. "Framing in Design: A Formal Analysis and Failure Modes." In *Proceedings of the 20th International Conference on Engineering Design (ICED15)*, July 27–30, 2015, Milan, Italy, Vol. 3: Design Organisation and Management, 133–142. Design Society.

Volti, Rudi. 1992. *Society and Technological Change*. 2nd ed. New York: St. Martin's Press.

CHAPTER 23

URBAN AESTHETICS AND TECHNOLOGY

SANNA LEHTINEN

1. Introduction

With postindustrial development of the past century, the role of urban environments has changed radically. The evolution of aesthetic interest in cities and urban life has followed a different trajectory in different parts of the world and also with cities of varying scale and status. Some great cities of global reputation such as Paris, Rio de Janeiro, or New York City have long been recognized for their significant landmark qualities and aesthetic value. The situation has been different for many other cities. In the United States, for example, disregard and neglect in the forms of suburban flight and racism affected how cities were perceived well until the 1990s and this is still reflected in many urban areas. Urban aesthetics takes other types of forms in cities globally, each embodying a multitude of distinct cultural contexts and approaches to urban design. Postindustrial development has dramatically reshaped the urban landscape and the values according to which cities are regarded and further planned. The steep increase in travel and tourism has played a role, as cities and their recognizable and unique features have become of interest to more people than ever before. The COVID-19 pandemic with the ensuing restrictions to the use of public space and travel has made these values more explicit and could further transform urban design values in the twenty-first century.

Concurrent with postindustrial development, a growing interest among late twentieth century urban planners in the human-scale quality of everyday life (Jacobs 1961; Gehl 1987) led to increased attention being paid to the types of experiences in which cities are able to engage their citizens. This attention has been further focused by grassroots level action such as tactical urbanism, urban activism, and place making. The practical interest is accompanied by a parallel development in the philosophical context, namely the acknowledgement of the situatedness of the human subject. As a result, much closer attention is paid in urban studies at large to how circumstances, place values, and physical locations affect subjective wellbeing while also having the potential to arouse imagination and curiosity. Cities are thus no longer considered a necessary evil, but instead central places for creating human wellbeing beyond economic values such as indicated by the GDP. This chapter aims at making explicit how, in connection to this new awareness, aesthetics is rarely if ever "only" aesthetics in the urban environment. The main goal is to show how the particular approach of aesthetic inquiry into perception, representation, and use of technology can enrich our understanding of urban life today and into the future.

Within different philosophical traditions, the value of the aesthetic approach has already been recognized well beyond the traditional sphere of the philosophy of art. Environmental aesthetics has explored the aesthetic value of natural environments since the 1960s and since the 1990s increasingly extended this exploration to human environments such as the built environment and the city (Lehtinen 2020b). Everyday life, which defines the use and planning of cities to a great extent, has been in the focus of everyday aesthetics since the early 2000s.[1] However, due to the original emphasis on the ecological health of natural environments, technology of any type has been a blind spot even when studying the aesthetics of such technologized environments as cities. The strength of a philosophical approach to urban aesthetics lies in how it can draw attention to the implicit values that are manifested in the city in terms of layout, artifacts, rhythms or how the city aesthetically conditions human relations and activity. It is worth noticing that the word 'aesthetic' does not equate with an interest only in beauty or positive aesthetic values in general, even though it is commonly used synonymously with these. Neither is the term 'aesthetic' used to describe only artistic or creative phenomena; following the first definition of the philosophical field by Alexander Gottlieb Baumgarten in 1735, it refers instead to the whole range of human sense-making that originates in perception.[2] Thus, everyday environments and the quality of living conditions are of crucial importance to aesthetic experience, as sense-making and aesthetic value processes are technologically mediated to a great degree in contemporary urbanized societies.

Walter Benjamin as one of the early cultural theorists had an aesthetically oriented approach to contemporary urban life. He gave emphasis to the Baudelairean figure of the detached *flâneur*, a wanderer, who observes and enjoys the city with an attitude of aloof interest. This usually privileged member (male, idle, intellectual) of society became an emblematic representative of the modern urban experience, observing the rapidly changing urban lifeworld.[3] Regarding fast-developing technologies of his era, Benjamin's account of the non-human disciplinary power of modern traffic signals revealed how "technology has subjected the human sensorium to a complex kind of training" (Benjamin 2007, 175). This already points to the development of the intersection of aesthetics and technology in studying the city, even though aesthetics is merely implied. In our contemporary societies, the aesthetic presence of technology is relatively easy to examine at the structural level of the urban lifeform, but it is nevertheless unclear what these different types of technologies signify for the deeper experiential processes of urban dwellers.

This chapter presents the links between urban aesthetics and technology as an emerging intersection of philosophical and applied aesthetics with the philosophy of technology. The focus is on analyzing the effects of technologies that are in use in contemporary cities from the perspective of aesthetics, with an emphasis on the urban experience at large. The theoretical background draws from environmental and urban aesthetics, which combines analytical, phenomenological, and pragmatic approaches, as well as a broad take on the philosophy of technology that combines pragmatist and postphenomenological approaches to new and emerging urban technologies. This framework is employed with practical applications in mind, such as urban planning or design. Combining urban aesthetics and the philosophical study of technology is thus a doubly specialized orientation, representing theoretical and pragmatic interest in a specific group of technologies, as well as adopting a specific viewpoint of aesthetics as its philosophical approach. Ethics also plays a significant role in this approach (as most times when any type of environmental aesthetics is discussed), as we shall see further in the chapter.

Interest in how the urban lifeform could and should be understood and developed further is not, of course, unique to philosophical urban aesthetics. Aesthetic and ethical considerations regarding cities have traditionally been presented in architectural or planning theory (Fox 2000). Many ideas about quality, livability, and aesthetically positive qualities in particular are being developed also within non-philosophical fields such as urban geography or ecology. 'Aesthetics' is often referred to directly in such studies, but an adequate definition for the concept is typically lacking. Moreover, communication across these highly specialized fields is still inadequate. The motivation of this chapter is to take this knowledge beyond disciplinary borders and to show in what ways philosophical aesthetics could be of use in reaching common definitions for shared concepts, as well as sharpening and directing future discussions on urban technologies, in theoretical as well as practical fields.

In addition, the chapter deepens a dialogue between aesthetics as a human experience-focused area of philosophy and city-oriented branches of philosophy of technology, while identifying the perspective of urban environmental aesthetics as a new direction for future research. Currently, research on aesthetic perception, values, and technology is scattered across different, often separately developing fields such as HCI, surveillance studies, digital aesthetics, and philosophy of the city. Philosophical urban aesthetics can provide a common ground for a broadly informed understanding of the deep impact of technology on human perception and aesthetic values in the urban sphere. Such an understanding is already long overdue, given the rapid development of technologies such as autonomous vehicles or drones, urban climate engineering, surveillance technologies, and increasingly refined augmented reality interfaces that will have a lasting impact on how cities look and feel. The links between aesthetics and technology in cities can be studied from two perspectives: the broader scale of the future-oriented approach to urban design and engineering, and the more nuanced and fragmented experiential knowledge of the inhabitants or other users of the city. A key insight in this chapter is that the first perspective is present and developed, for example, in smart city discourse and practices, whereas the latter does not yet receive enough attention.

2. Aesthetics and the Technologically Mediated Urban Experience

On a general level, a *descriptive* urban aesthetics focused on technology charts what types of technologies are in use in cities and how they alter the perception and aesthetic appreciation of an urban space or place. The *normative* interest in urban aesthetics goes beyond the mere descriptive and perceptual layer, to assess the reasons for cities or their elements looking and feeling the way they do.[4] In the particular case of urban technologies, normative aesthetics would thus aim at explicating the conditions that would maximize the positive effects and minimize the adverse effects of technology on the aesthetic experience at large of those living in urban areas. This includes not only the effects of planned features, but also of the unplanned, unintended, and non-human experiential elements. A critical and evaluative frame is thus implied by the normative approach, whereas the descriptive approach is commonly used in representations of the city ranging from travel stories and advertisements to literature, visual art and popular entertainment. In this chapter aesthetics refers to a broader aesthetico-ethical framework for *assessing* perceptual factors. However, this section will first lay out some conceptual frames that have been useful in articulating the domain of urban aesthetics and the ways that technologies mediate our experience of the city.

Besides the descriptive/normative distinction, it is useful to distinguish between the two main layers in urban aesthetics more deeply, as this helps to clarify some confusions regarding the use of the word 'aesthetic.' The division between the surface and the deeper layer of aesthetics is based on the recognition of the "thick sense" and the "thin sense" of aesthetic experience, which Allen Carlson has applied to the layers of aesthetic appreciation in environmental aesthetics (Carlson 1976, 75; Carlson 2005, 142; Carlson 2009, 94–95).[5] In the case of an urban technological artifact judged aesthetically, the thin sense refers to appreciating the formal features that express and convey the design outwardly. This might refer to basic visual features such as form, color, composition, and the rhythm of its motion for example, or the way in which it either blends in or stands out in the cityscape. As an example, we could assess what type of impact different types of urban mobility technologies such as trams or electric scooters have for the urban landscape. The thick sense refers to further assessing the aesthetic impact of these artifacts, based on how well their use reflects the more implicit values of its design. With the example of urban mobility technologies, this could mean taking into account the ecological or social impacts of the technologies in question. Other values are not necessarily to be treated as direct trade-offs with aesthetic value, but recognizing the broader value implications of the technologies in question has an impact on how the aesthetic value is constructed in the urban sphere.[6] In the process other values "seep" into the aesthetic experience and ensuing judgment. This includes recognizing how the artifact or system responds to ethical, environmental or functional needs and in which way this connects to its aesthetic impact.

The "thin" or surface level aesthetic appeal is in focus in the *macro layer of urban aesthetics* as a more detailed approach (Lehtinen 2020b). I am using it to refer to the impactful, recognizable and often visually perceived features of a city. A high-rise building, for example, is as such a distinct form and also a collection of advanced technologies that has a deep impact on the overall look of a city. The macro layer of urban aesthetics is very much present in the common manner of representing and conceptualizing cities, from tourist brochures and cities as movie backdrops to architectural renderings and city branding.[7] The aesthetically prominent features such as large-scale buildings, monuments, and pieces of large infrastructure and the overall "look" of the city are of interest especially in the world-famous metropolises that are defined by instantly recognizable unique features. However, it is clear that this focus on surface or even relational aesthetic qualities is not the only way to evaluate the aesthetic impacts of technology in the urban sphere.

The "thick" sense of aesthetics is more pronounced in the so-called *micro level of urban aesthetics* (Lehtinen 2020b). Going beyond most obvious urban imageries, it zooms into the subjective experience on an embodied, engaged, and sensorial level.[8] This deeper interest in how aesthetic perception constructs us as subjects, describes the experience of the "everyday life flow of the city with its inconsistencies, contradictions and messy relationships" (Shane 2002, 235). This comes closer to the phenomenological approaches to the aesthetic experience, as it enables focusing on the subjective and perceptual elements of interaction with the world. The idea of 'engagement,' with roots in pragmatist philosophy, is central to this deeper layer of the aesthetic experience (Berleant 2010). The distinction between the consistent, even monotonous everyday urban experience and more transitory or highlighted ways of engaging with the city is significant for recognizing the preconditions of the experience such as attention, attitude, expectations, and biases.[9] Following the thinking made explicit by the notion of thick sense, our experiences and perceptions in situated interactions with technologies within the urban sphere are hardly ever devoid of ulterior meanings or beyond the scope of concerted effort to understand their implications. Instead, these moments reveal the prevailing values precisely through what it is that we are seeing, hearing, smelling, feeling, or perceiving in other ways, and in which ways we are responding and reacting to those perceptions.

Instead of attending only to the visual features of the urban environment, the inevitable multisensoriness of the urban aesthetic experience is to be recognized. The different sensory modalities of sight, touch, sound, smell, and taste and their intricate interplay are central when the city is experienced through daily habitual interactions with and within it. An interesting example of changes at this level is how, with new and emerging technologies, the sense of touch has become increasingly central through haptic and gesture-based interfaces in the built environment. Besides the sensory modalities it intrigues, the scale of technology can be an important factor in determining its aesthetic impact. One useful way to approach urban technologies is through postphenomenological analyses of their use. This approach recognizes the value-laden and value-forming position of technologies between humans and the world, but emphasizes that they work in ways which end up changing both their users and the world between which they are functioning (Ihde 1990; Verbeek 2005). From this perspective, technologies become perceptually and epistemologically important insofar as the world becomes interpreted as well as operated through them. Also, importantly for the thick sense of aesthetics, the postphenomenological approach emphasizes that technologies are never neutral (Verbeek 2005). They are from the beginning designed, built and invested in with certain prevailing values and goals in mind. Even though these technologies might end up being used in myriad ways, their design still carries a set of the originally limiting ideas and values.

The postphenomenological approach to technological mediation has been studied in more detail in the case of urban design, especially on the level of more or less movable objects. Robert Rosenberger, for example, starts with a postphenomenological analysis of the *multistability*[10] of urban objects and proceeds to analyze the ways in which hostile and oppressive ideology toward certain groups of people (e.g. the homeless) is designed into their experiential and perceptual features (Rosenberger 2017; Rosenberger 2019). Examples include park benches with built-in dividers to stop people reclining or resting, ledges with obtrusive spikes, or trash bins with sealed covers to stop people collecting bottles. This type of hostility is also directed toward other marginalized groups of people (anti-loitering design)[11] and to an even greater extent toward non-human species considered unwanted in the urban space (e.g. pigeons, rats), generally accepted unquestioningly for security and sanitation reasons. These "built-in" hostile features of urban environments might not be perceived by those not directly affected by them, but once one realizes the logic of this type of design, one can no longer "unsee" the hostile design features. The process of *aesthetic disillusionment* can be used to describe how the recognition of the hostile intentions of a technological design causes permanent friction with how the object is perceived and evaluated aesthetically.[12]

The idea of a kind of "innocent" perception is still quite strongly linked to how aesthetic appreciation is generally understood. This is partly due to the long tradition of emphasis on "disinterestedness" as a defining factor of a true aesthetic judgment.[13] In order to be considered genuine and authentic, the aesthetic judgment has been ascribed the need to detach from other motivations one might have toward the object of evaluation. This idea was solidified by Immanuel Kant in the *Critique of Judgment* (2000 [1790]), but before this canonical structuring of Western philosophical aesthetics there seemed to be more flexibility in understanding the complexity of human aesthetic motivations and intentions.[14] Disinterestedness is a debated principle even in philosophy of art (where it has been used as, for example, a defense for the autonomy of art), but it is clear that especially in relation to urban environments, disinterestedness of aesthetic judgment becomes re-evaluated due to the very practical nature and use of these environments and the human and beyond human (e.g. technological) elements of which they consist. Where aesthetic and practical interests clearly coexist and overlap, the concept of disinterestedness becomes relevant in determining to *what extent* other values affect aesthetic judgments more than *whether* they do so. This logic is presented by the so-called cognitive approaches in environmental aesthetics which discuss the role of scientific knowledge in the aesthetic appreciation and evaluation of natural environments (Carlson 2019). Non-cognitive approaches in environmental aesthetics complement this by reminding us that other elements of the human experience are also present in aesthetic encounters with the environment, for example imagination (Brady 1998), a sense of mystery (Godlovitch 1994), memory, and human emotions. All these various factors modulate the extent to which aesthetic judgment is affected by factors other than the "purely" aesthetic ones (Carlson 2019). The non-cognitive approaches emphasize embodied and immersive experience instead of informed and distanced judgment, pointing toward how perception and aesthetic appreciation take place in a complex, temporally and even spatially evolving process.

Cognitive approaches in aesthetics of natural environments have been developed based on increasing ecological awareness and concern, but the same logic is applicable to the relation of perceptual experience and cognitive factors in technologized urban environments. A broad spectrum of human capabilities affects aesthetic judgments: besides rational thinking and scientific knowledge, aesthetic impressions, judgments, and preferences are also heavily influenced by subjective factors such as previous experiences, emotions, and imagination which, despite their differing contents, are common to all humans. The dimensions of this process become especially clear in the everyday environments that one perceives through the haze of familiarity and routine behavior (Haapala 2005). The subjective factors affect intuitive reactions and attitudes that govern to which phenomena one pays attention. Whenever strong subjective stances are recognized as such, they often reveal interesting biases in our aesthetic appreciations. For example, we might be prone to nostalgia for the aesthetic trends of our childhood, all the while recognizing that those trends are based on ecologically unsustainable choices. The cognitive and non-cognitive approaches and their relation to the surface and deeper levels of aesthetic appreciation help to differentiate between intuitive and more elaborated judgments about what one perceives in the urban sphere.

In the context of aesthetics of urban technologies it is worth thinking further if 'interesting' or 'noteworthy' have to some extent replaced traditional positive aesthetic values such as 'the beautiful,' as commonly used aesthetic characterizations. This implies, that what catches the *attention*, even by such negatively tinted aesthetic qualities as ugliness or messiness, is considered more valuable than the things that are less noticeable. As an extreme form, this is visible in the attention and experience economy which shows how the value of heightened experiences has become more central in postindustrial societies (Pine and Gilmore 1999). However, the propensity of aesthetic phenomena to draw attention and raise interest can be a potential signifier of other important values. According to John Hospers, paraphrased in environmental aesthetics by Carlson, objects are expressive of fundamental “life values” in the thick aesthetic sense (Carlson 1976, 75). According to this line of thinking, we pay attention to and have a strong preference for aesthetic features in the environment that support our purposes beyond the merely pleasant or visually beautiful (Berleant 1992; Besson 2017). This attention and subsequent preference thesis have important implications for the evaluation and appreciation of new and emerging technologies, as we shall see further on in this chapter.

Another distinctive feature of the aesthetic experience of technology in the urban sphere noted in both sociological (Simmel 1969) and philosophical (Welsch 1991) literature is a kind of experiential numbness or indifference as an “anaesthetic” quality, one that has been used to describe large-scale metropolises. For the sociologist Georg Simmel, this experiential numbness follows from the overload of the human perceptual capacities caused by “the *intensification of nervous stimulation*” which, for him, is part of the broader oppression of the socio-technological mechanism (Simmel 1969, 47–48). According to this view, human experience in stimuli-filled modern cities becomes a numbing condition which further alienates one from more authentic, vivid and enjoyable experiences in life. This represents a broader theme in contemporary scholarship, that of caution both toward technology and contemporary forms of urban and societal life. Opposed to this pessimistic view is the frame of technology as enabler of a new, heightened level of *positive* intensification, precisely through its mediating of the urban aesthetic experience. The diverse metropolitan cityscapes of East Asia, for example, have become defined by technologies such as flashing neon and LED screens that are seen as making them culturally (and not only, e.g., economically) significant; such cities are widely appreciated for their vibrant and rich display of aesthetic stimuli, well beyond what was considered desirable or even healthy just a few decades ago.

Finally, technology is also linked to the developing notion of the urban sublime. The concept of the ‘technological sublime’ has entered the aesthetic discourse and vocabulary to describe postindustrial nostalgia toward traces of past human activity in the cultural landscapes (Nye 1994). The sublime in general, and in the urban context in particular, has been used to give access to the experience of limitlessness and being part of something bigger in the crowds of the city: of how the city enables an experience of a “multitudinous humanity” (Den Tandt 1994; Den Tandt 2014). The use of technology has the potential to accentuate or suppress different types of narratives in the urban sphere, and elements of surprise and an imminent sense of danger build up the sublime experience (Berleant 2007). Faithful to the previous formulations of the concept by Edmund Burke and Kant, the sublime is also characterized by awe-inspiring spatial dimensions, for example as found in high-rise buildings or entire skylines (Nye 2005). As the use of technology is often linked to increased safety measures in cities, it also has a direct link to situations that have the potential to be experienced as sublime. Enjoying the sublime aesthetic qualities of previously dangerous phenomena such as earthquakes is at least theoretically possible, if one’s life is not directly under threat thanks to modern earthquake detection systems and earthquake-proof technologies in the built environment.

3. Further Aesthetic Implications of Urban Technologies

The influence of technology on urban aesthetics can be studied from two further perspectives. The first charts how the look of urban environments is changing due to entirely new technological innovations. This perspective emphasizes the disruptive nature of new and emerging urban technologies and is commonly used in smart city discourse, when presenting entirely new state-of-the-art technologized neighborhoods, for example. The other, more nuanced point of view relates to how new and incrementally developing technologies mediate the experience of already existing and historically layered urban environments. The first perspective is often present when new technological innovations are being pitched and the latter when discussing retrofitting or upgrading infrastructural technologies, for example. In comparison with the broader conceptual frames outlined in the previous section, these discourses direct our attention to the urban aesthetic impacts of individual technologies instead of technology writ large. This follows also the 'empirical turn' in the philosophy of technology:[15] in practice, it might not make much sense to discuss the aesthetic consequences of traffic lights and the 5G network in the same context.

The concept of affordance has been used to describe how particular technological artifacts function and what type of activity they make possible in the city, for example, the aesthetic and ethical implications of bridges as large-scale infrastructural artifacts (Winner 1980; Allen 2008). Borrowed from ecological psychology and applied to the social study of technology, the concept of an affordance helps to explain how perceptual features are interpreted in experience (Gibson 1979; Bloomfield, Latham, and Vurdubakis 2010). Affordance also helps us to understand the specific ways in which particular technologies affect the urban experience. Broadening (but also sometimes limiting) the affordances at hand, technologies give new meanings to objects that previously were perceived with different ends in mind. Perceiving affordances as action possibilities takes place in a socio-culturally conditioned and temporally definable moment (Vihanninjoki 2020; Lehtinen and Vihanninjoki 2021). From our earlier glimpse at hostile design as an example, it is clear that technological objects can also have political affordances. They often embody and reify societal hierarchies of power, especially in the urban sphere (Winner 1980; Rosenberger 2017). The sensorial features of the design of those objects also have the potential of making these power relations explicit. However, as we become accustomed to the presence and uses of these technologies, these links start to become opaque to us.

Intuitively, it does not make much sense to discuss the aesthetics of technology without taking into consideration functionality and how well the designed technology fits the purpose for which it is designed. This follows the idea of design of any type as the field of *dependent beauty,* as opposed to *free beauty* present in art or nature (Forsey 2013, 137–140; Parsons and Carlson 2008). Linking aesthetics and functionality becomes increasingly complex, when one recognizes that technological designs also follow aesthetic trends or presuppositions regarding their aesthetic features. In architectural design this is already recognized in the phenomenon of modernism, which very soon after setting its original functionalist goals, created an aesthetic style that sometimes even contradicted the functionality of its design features (Schummer, MacLennan, and Taylor 2009, 1038).

A technology can draw attention through its aesthetic qualities either intentionally or unintentionally. When intentional, this is usually planned in the design process. In contemporary cities, technologies can also be used to draw attention to different, already existing aesthetic features of the environment, making detectable something that previously went literally unseen. Architectural details, for example, which have faded into the background for centuries might be brought to the focus of attention by a new façade lighting system or when AR games such as Pokémon Go feature them. Converging digital and infrastructural technologies thus alter human perceptions of the city (Caldwell, Smith and Clift 2016). The politics and aesthetics of information technologies in particular have been studied through the concept of the interface. This encompasses a range of technological designs such as displays, screens and smart devices in an attempt to study their role as mediating thresholds that generate processes of interaction in the city (Galloway 2012, vii; 121). The interface approach focuses on digital media technologies and frames them as aesthetic (or even poetic) objects and, as such, aesthetically relevant (Galloway 2012; Manovich 2001).

Traces of technologies in cities are omnipresent. As urban technologies are becoming increasingly less object-based and instead more convergent and networked, recognizing them as technologies in the first place is less straightforward. Yet such networks often inject new rhythms into urban life, as exhibited by Benjamin's account of the impact of traffic signals or in analyses of the smart city development (Picon 2015, 138). Importantly, technology also adds a new layer of care into the urban environment. Care through maintenance is becoming an increasingly central theoretical framework for understanding the human concern for objects and environments (Mattern 2018; Saito 2020; Lehtinen 2020b). Urban technologies deteriorate, solutions become outdated, infrastructures crumble, and overseeing their maintenance and making decisions about these changes depends on active human agency. Aesthetically manifesting features such as rust, dirt, and decay are indicators of the material condition of many technologies, even though less present in assessing the conditions of technologies that "hide in plain sight," such as the water supply or Wi-Fi networks.

In environmentally attuned aesthetics, cities are discussed in terms of the notion of 'authenticity.'[16] Instead of the origins of the city, authenticity is used in this context to describe environments which allow people to flourish. Lack of authenticity or even straight-forward falseness, according to this approach, equates to an intolerable degree of compromise for human wellbeing, for example in situations when the city might work technically or economically well but does not support other, more humane values in its design (Berleant 1992; Besson 2017). Technologies that help cultivate aesthetic sensibility and humane engagement with the city might thus be especially well-designed to foster urban authenticity: in a similar way as technologies can guide attention to previously underacknowledged aesthetic features of the city (Lehtinen and Vihanninjoki 2019), they can facilitate new attitudes and skills among their users.

4. Aesthetics and Urban Mobility

The development and implementation of many prominent technologies during the past 150 years has changed and developed cities widely. It is well recognized how automobiles, for example, affected urban design and the patterns of everyday life in fast-growing Northern American cities. Transportation safety has been a defining factor for the design of spaces for motorized vehicles for obvious reasons. However, how safety is interpreted and transcribed into the aesthetic features of the urban environment has erred on the side of caution even to the extent of the creation of homogenous, sterile streetscapes with no fixed risky elements such as trees or other things that might distract the attention of drivers (Dumbaugh and Gattis 2005). Increasing safety is also one of the goals of artificial illumination such as streetlights in urban environments.[17] However, it can have serious harmful effects on the ecosystem level, especially to non-human species (Stone 2017; Stone 2018). In addition, artificial lighting contributes to the increase in the time during which human activity is possible, leading to societal expectations of 24-hour cities (Adams et al. 2007).

The history of urban transportation has many good examples of large-scale changes made in the name of progress and adding efficiency to the use of the city. The aesthetic achievements of these changes have not gone unnoticed, for example when the "technological beauty" of the railroads was acknowledged (Rice 1997, 208). Technology in these cases is read aesthetically as a sign of human intelligence, progress and new opportunities. Technology also gives form to the historical traces and residues of human activity. Industrial cityscapes are a good example of this. Thus, besides aesthetic features, knowledge about the content and the context are needed for interpretation—this might include functional, historical, relational features, and so on (Stolnitz 1960). Such acts of aesthetic interpretation range from intuitive to more complex processes of meaning-construction.

Mobility is an area of aesthetic inquiry that has been discussed in relation to urban environments to the extent that even the term "mobile aesthetics" has been used (Naukkarinen 2005). How humans move, especially in urbanized areas, is not an insignificant factor for the experience of the city or the overall quality of life. Different transportation modalities such as walking, bicycle, car, bus, metro, or tram integrate technology and have a high impact on lived experience. The city looks different when using different modalities. Different aspects of the city are perceived from the bus to those perceived when cycling, for example. The interest in the figure of the *flâneur*, a leisurely urban stroller, has persisted, since walking as a kinaesthetic modality of movement binds together a flow of impressions of the city temporally and spatially (Lobo

2020). Walking interfaces with the city using a minimal amount of technologies (unless the walking experience is mediated by smartwatches, collection of locational data, headphones, etc.), but it also relies on technologies such as traffic lights in order to communicate with other users of the road infrastructure.

With the development of autonomous or self-driving vehicle (SDV) technologies, the unpredictable element of driver behavior might no longer be part of the safety equation. Here, instead, the human agency involved in driving the vehicle is almost entirely replaced by smart technologies within the vehicle itself (e.g. LiDAR) as well as embedded in the environment (sensors). This does not, however, take away the human factor of those using other modes of transport, such as pedestrians and cyclists. Valid fears do exist, that the shift to autonomous vehicles will lead to undesirable changes in urban design since solving the risks to safety and eliminating uncertainty of human behavior is not as straightforward as at first assumed. This could mean in practice adding fences and rails to deter pedestrians from taking shortcuts, or adding obtrusive sensory cues such as sound signals to pedestrians and cyclists in order to increase their safety.

Urban mobility in general is a good example of a human system and practice that is reflected in the micro layer of urban aesthetics; how it affects the everyday choice of routes and transportation modalities by determining what is possible in the first place. Urban mobility has not, however, been extensively studied from the perspective of how it either produces or hinders subjective aesthetic experiences in everyday urban life (Mladenovic et al. 2019). The impact of driverless cars on how the city will be used has been difficult to assess from the perspective of social justice or its effects on vulnerable people (Epting 2019). Aesthetically speaking, this concerns questions such as who has the right to beautiful scenery, pleasant surroundings, or stimulating open vistas along their transportation routes. This makes explicit how aesthetic values are, once again, accompanied by complex ethical considerations and lead to prioritizing between not only values but also groups of users of different transportation modalities.

New mobility technologies can also emphasize traditional scenic beauty in cities, a good example of this being the ubiquitous dronescape photography and video. It has been preceded by aerial photography and outdoor photography in general,[18] but the new techniques and relatively easy accessibility of drone cameras have vastly expanded the number and style of images of cities taken from above. With their use ranging from tourism marketing material to real estate business, such photos and drone videos have become so common that they are affecting how the city is envisioned and how its spatial conditions are perceived and critically studied (Jensen 2020). The aerial perspective is not available to humans without the significant extension of technology. These and many other types of technology-enhanced urban experiences are becoming increasingly central to how the city is perceived.[19]

Digitalization of the urban sphere has been designed and studied in recent years with experiential and aesthetic affordances in mind (e.g. Andersen and Pold 2018; Caldwell, Smith and Clift 2016). Virtual reality and AR solutions augmenting the sensory realm or GPS-based location applications, such as wayfinding tools, for example, potentially broaden the scope of aesthetic experiences (Lehtinen and Vihanninjoki 2019). As a form of experimental aesthetics in the urban sphere, GPS technology has also jump-started a phenomenon of locative media arts; the work of artists such as Masaki Fujihata, Teri Rueb and Christian Nold helps to make transparent how the patterns and trajectories of urban dwellers are traced and recorded. Locative media art makes the functioning of technologies visible and play with visibility and invisibility of their effects. These artworks underline the broader friction between the invisibility of many new and emerging technologies and the perceivable traces of their use. GPS and 5G technologies in themselves, for example, are perceptually indiscernible, but nonetheless enable bringing into use a group of other technologies (e.g. SDVs require widespread implementation of 5G networks) which will change significantly the look, use and experience of cities. Another type of invisible technology are environmental sensors, for example, for detecting airborne particles that indicate poor air quality. Such harms might have been already perceptually detectable for a long time in the form of smell or visible smog, but technology adds a layer to this sensory apprehension, as Hanna Husberg and Agata Marzecova have shown in their collaborative work between art and ecology.[20] Urban environmental harms receive the attention they require often only once the perceptual accounts of them are accompanied by scientific proven and technologically measured results.

5. The Relevance of the Aesthetics of Technology for Urban Design

Urban planning aspires to make the city a better place in terms of its many functions and the life of its inhabitants. Interpretations of what this means include increases in safety, efficiency, and affluence, for example. In this sense, thinking about urban futures is bound to be a utopic and speculative endeavor. This is not only because ideals and generations change but also because urban technologies evolve and become obsolete. Finding one, unanimous goal for urban planning is thus also utopic; urban design and planning is better understood as an ongoing, multi-dimensional process with a focus on making compromises. The current interest in participatory methodologies in urban planning and academic studies implies a move toward more inclusive planning and multidimensional understanding of the experiential qualities of cities. Subjective accounts of experiences and “opinions” about aesthetic matters can offer a valuable route to understanding the layers of values that incarnate in everyday interaction with the city. The traditional criteria for beauty as a positive aesthetic value have included qualities such as harmony, graceful proportions, or a moderate amount of diversity. When developing urban communities, attention needs to be given to a broader set of criteria. The details of the experiences might vary, but an overall positive quality of a living environment is linked to how it cultivates curiosity and fascination. This is possible if cities offer "variety, challenges, and even negative experiences, not eternal bliss and ease" (Besson 2017, ch. 5).

The methodology of taking aesthetic experiences into account is still being developed even in the most participatory approaches in urban design and planning practices. The sphere of aesthetic experience is still commonly linked only to individualistic or hedonistic pleasure detached from everyday matters. As I have noted elsewhere, this surface-oriented understanding of aesthetic experiences might seriously hinder understanding the ethical implications of the aesthetic realm (Lehtinen, 2020a). It also implies that experiences can be controlled and designed to a greater extent than is most likely possible, as aesthetic values show a tendency to change with time. Another issue related to this is the scope of urban planning and aesthetics. Urban infrastructure and individual buildings form the largest visible part of our involuntarily inherited legacy. What happens *between* buildings is equally important, if not more so, from the standpoint of urban aesthetics.[21] In human-dominated environments, these in-between spaces are often overlooked with their features not considered to be central to the overall urban experience. Meanwhile, as the case of mobile aesthetics demonstrates, such an everyday phenomenon as urban transportation consists mainly of networks, routes, and places which have nothing to do with the "main attractions" of the city.

As implied already earlier in this chapter, the experience of the urban everyday landscape is also affected by urban design strategies that employ technologies which have security as their main goal. Counter-terrorism security measures are thus one clear group of technological designs that have aesthetic consequences for the urban environment, but which citizens usually do not have the opportunity to decline, or might not even recognize their function (Coaffee, O'Hare and Hawkesworth 2009). Besides clearly obtrusive security elements such as fences, even more nuanced security measures such as anti-homeless or otherwise inaccessible design features increase exclusionary experiences in urban areas and make it clear that some individuals and groups of people are more welcome than others. The design of the urban landscape thus always reflects what security policies the city governance is relying on. This is of course related to the phenomenon of hostile or defensive design presented earlier, which aims on a smaller scale to keep unwanted social groups or non-human species away from public spaces.

In cities, current generations are living with discernible signs of the values and activity of the preceding generations. In the same way, the decisions made today affect the aesthetic range of experiences for future generations (Lehtinen 2020a). As updating elements of the built environment affects how they are perceived and used (Bouzarovski 2015), implementing new technologies requires acknowledging this intergenerational temporal dimension.[22] Transgressing aesthetic codes or perceived "normalcy" is often used as a strategy for introducing new technological elements into the urban sphere, although it has also its risks. An example is the increase in digital screens with moving images used mainly for advertising but sometimes also for cultural and entertainment purposes. The screens are highly visible and as such often disliked at first due to the bright lights and attention-grabbing movement (Kolhonen 2005). However, after the initial phase of their introduction many are quietly accepted and rarely paid involuntary attention to anymore. As this example shows, the experience of urban technology is complex, as it is often uncertain whether a technology will increase the occurrence of the experience of beauty or end up hindering it. New urban technologies negotiate and test the boundaries of aesthetic experience.

A specific area of urban aesthetic interest, where art and technology come together, encompasses special events and spectacles that involve technology. These, as all artistic forms, encourage us to see a familiar environment in a new light, even quite literally. Forms of art and entertainment with collective and participatory functions, such as carnivals, urban festivals, and other open-air gatherings, make use of urban space beyond its most clearly practical functions (Browne, Frost, and Lucas 2018). These types of festivities represent the extraordinary aesthetic dimension of urban public life (Leddy 2012). The city is always more than just a functional or aesthetic space but through these important elements it is possible to gain a more comprehensive picture of it. The traditionally aesthetic sphere of the human activity of art is linked to the ongoing exploration of technological mediation in contemporary urban space. In cities, as elsewhere, artistic forays into various technological media provide additional ways for new technologies to become naturalized and accepted. Parallel processes of technologization of the city space through aestheticization take place via advertisements, games and so forth. This is one of many reasons why the context of technologically mediated art can be a valuable source of insight informing urban design practices.

6. Conclusions

From the start, technology has been deeply embedded in the building stock, infrastructure, mobility, and indeed all facets of a city's activity. The current smart city ideology, as well the ecology-driven urban sustainability framework, both rely strongly on emerging technologies such as smart mobility or runoff water management systems which affect greatly how people experience the city. The standards of living and growing environmental challenges of contemporary societies globally have precipitated reliance on an even faster pace of development of many entirely new types of technologies.

Urban aesthetics studies how different types of aesthetic values manifest in urban environments and whether and how conflicts in values are resolved in them. As we move further into the twenty-first century, the aesthetic identity of different types of cities is changing due to new and emerging technologies. Examples such as mobility-related technologies show that urban aesthetics is not only a question of design or making things attractive, but instead involves a more fundamental part of human meaning-making in urban settings. Cities also offer a particularly aesthetic richness and diversity. A central insight in this chapter has been that aesthetic factors play an important part in our interactions with a broad variety, if not all, types of urban technologies. Another key aim has been to show, that aesthetics is rarely, if ever, only about the surface qualities of cities. Instead, even the obvious aesthetic qualities bear the potential of making other values visible. In this way, perceptual features of technology are key components of drawing our attention to certain functions and values, in the same way as the omittance of certain features is a sign of suppressing other values. Design choices signify deeper commitments to human values. In this sense, what is perceptually missing from the urban sphere is equally important to what is present. Philosophical approaches to the perception and the use of technology should thus be reframed with this in mind.

This chapter has observed the links between aesthetics and technology in cities from two perspectives: that of top-down urban design and engineering and that of the inhabitants or the users of the city who dwell in the everyday experiences of urban technologies. The aim has been to show how an aesthetic approach to urban technological phenomena is not reducible to the descriptive perspective, but encompasses also the reactions, habits, values and norms embedded in interactions with and in the city. Many theories in urban aesthetics are applied from environmental aesthetics which often borrows concepts and logic from the philosophy of art. Beyond these approaches, new ways to ask questions and to see the role of the aesthetic in relation to human activity are needed, as the sphere of aesthetic considerations can challenge habitual ways of being, sensing, and doing.

This means also that the aesthetic in the urban environment is always intertwined with the ethical. This also applies to the forms that technology takes. Just as ethical factors have to be taken into account when assessing aesthetics, aesthetics as such is important in bringing ethical issues to light. As the chapter has shown in the cases of hostile design and urban mobility development, socially unjust decisions of the past are made explicit as physical forms and aesthetic affordances in the urban sphere. Layered on to these dubious legacies of past generations of decision-makers are newer types of elements in the urban sphere (for example, surveillance technologies) which are changing the urban sensorium in ways that often go unnoticed. A thorough and continuous assessment of the multidimensional qualities of urban environments thus requires both aesthetic sensibility and methodology, so that we may evaluate and rethink what there is to be perceived and experienced, why, and in what ways.

Notes

1. For the aesthetics of natural environments, see Carlson 1976; Carlson 2009; Carlson 2019; for aesthetics of human environments, see Berleant and Carlson 2007; Berleant 2010; for aesthetics of the everyday, see Light and Smith 2005; Saito 2007; Leddy 2012; Haapala 2005; Haapala 2017; for urban aesthetics, see Lehtinen 2020b.
2. *Epistêmê aisthetikê* meaning the science of what is sensed and imagined, Baumgarten 1983 [1735], 86–87.
3. For the concept of the 'lifeworld' in the urban context, see Madsen and Plunz 2002.
4. For the distinction between descriptive and normative approaches in environmental aesthetics, see Berleant 1992.
5. The distinction between thick and thin aesthetics was developed first by D.W. Prall and John Hospers in the field of philosophy of art (Carlson 2005, 142).

6. This type of understanding of the relation between different values has been pronouncedly present in environmental and everyday aesthetics, although the paradigm of disinterestedness in philosophical aesthetics is still strong in discussions concerning the sphere of art.
7. For a characterization of the 'tourist gaze,' see Urry and Larsen 2011.
8. For *somaesthetic* accounts of the urban aesthetic experience, see Shusterman 2019. For embodied *kinaesthetic* urban aesthetic experience, see Lobo 2020.
9. Lehtinen 2015. For philosophical everyday aesthetics, see e.g. Light and Smith 2005; Saito 2007; Leddy 2012.
10. For the concept of 'selective permeability' in similar use, see Crippen and Klement 2020.
11. E.g., for spatially "planning out" teenagers, see Pyyry and Tani 2016.
12. *Aesthetic disillusionment* has been described in environmental aesthetics most notably by Cheryl Foster, who (echoing a recognition by Immanuel Kant) points out one can no longer admire the beautiful red colors of a sunset without the gnawing recognition of the possibility of human pollution being the cause behind the strikingly beautiful colors on display.
13. For the Kantian origin and development of the concept of disinterestedness specifically from the perspective of environmental aesthetics, see Carlson 2019.
14. E.g. in how David Hume has approached aesthetics through the notion of taste, see Hume 1985 [1757].
15. For an article that brings aesthetics together with design and user perspectives on technology, see Cammers-Goodwin and Nagenborg 2020.
16. See e.g. Vihanninjoki 2019; Wittingslow 2021.
17. For aesthetic appreciation of urban darkness, see Tainio 2019.
18. For the early days of aerial photography and its influence on urban planning and the urban experience in nineteenth-century Paris, see Rice 1997.
19. The COVID-19 pandemic, for example, has quickly led into imagining situations in which software applications for digital contact tracing, button-free elevators, and robots sanitizing the surfaces of buildings are part of the normal urban experiential realm.
20. Husberg and Marzecova's project: "As Air Became This Number," https://as-air-became-this-number.schloss-post.com/number.html
21. For the aesthetics and politics of the in-between urban spaces, see Mubi Brighenti 2013.
22. Retrofitting describes the process of adding new technology or new features to older, previously existing systems, even though there is currently no consensus on the exact definition, see Eames et al. 2018.

References

Adams, Mags, Gemma Moore, Trevor Cox, Ben Croxford, Mohamed Refaee, and Steve Sharples. 2007. "The 24-hour City: Residents' Sensorial Experiences." *The Senses and Society* 2, no. 2: 201–215.

Allen, Barry. 2008. *Artifice and Design: Art and Technology in Human Experience*. Ithaca NY: Cornell University Press.

Andersen, Christian Ulrik, and Søren Bro Pold. 2018. *The Metainterface: The Art of Platforms, Cities, and Clouds*. Cambridge MA: MIT Press.

Baumgarten, Alexander Gottlieb. 1983 [1735]. *Meditationes philosophicae de nonnullis ad poema pertinentibus/Philosophische Betrachtungen* über *einige Bedingungen des Gedichtes*, edited by Heinz Paetzold. Hamburg: Felix Meiner Verlag.

Benjamin, Walter. 2007. *Illuminations*. Edited by Hannah Arendt. Translation by Harry Zohn. New York: Schocken Books.

Berleant, Arnold. 1992. *The Aesthetics of Environment*. Philadelphia: Temple University Press.

Berleant, Arnold. 2007. "Cultivating an Urban Aesthetic." In *The Aesthetics of Human Environments*, edited by Arnold Berleant and Allen Carlson, 79–91. Peterborough: Broadview Press.

Berleant, Arnold. 2010. *Sensibility and Sense: The Aesthetic Transformation of the Human World*. Exeter: Imprint Academic.

Besson, Anu. 2017. "Building a Paradise? On the Quest for the Optimal Human Habitat." *Contemporary Aesthetics* 15. https://www.contempaesthetics.org/newvolume/pages/article.php?articleID=806

Bloomfield, Brian P., Yvonne Latham, and Theo Vurdubakis. 2010. "Bodies, Technologies and Action Possibilities: When Is an Affordance?" *Sociology* 44, no. 3: 415–433.

Bouzarovski, Stefan. 2015. *Retrofitting the City. Residential Flexibility, Resilience and the Built Environment*. London: I. B. Tauris.

Brady, Emily. 1998. "The City in Aesthetic Imagination," in *The City as Cultural Metaphor: Studies in Urban Aesthetics*, edited by Arto Haapala, 78–92. Lahti: International Institute of Applied Aesthetics.

Browne, Jemma, Christian Frost, and Ray Lucas, ed. 2018. *Architecture, Festival and the City*. London and New York: Routledge.

Caldwell, Glenda Amayo, Carl H. Smith, and Edward Montgomery Clift, eds. 2016. *Digital Futures and the City of Today. New Technologies and Physical Spaces*. Chicago: Intellect, Chicago University Press.

Cammers-Goodwin, Sage, and Michael Nagenborg. 2020. "From Footsteps to Data to Art: Seeing (through) a Bridge." *Contemporary Aesthetics*, Special Vol. 8. https://contempaesthetics.org/2020/07/16/from-footsteps-to-data-to-art-seeing-through-a-bridge.

Carlson, Allen. 1976. "Environmental Aesthetics and the Dilemma of Aesthetic Education." In *The Journal of Aesthetic Education* 10, no. 2: 69–82.

Carlson, Allen. 2005. *Aesthetics and the Environment: The Appreciation of Nature, Art and Architecture*. London and New York: Routledge.

Carlson, Allen. 2009. *Nature and Landscape: An Introduction to Environmental Aesthetics*. New York: Columbia University Press.

Carlson, Allen. 2019. "Environmental Aesthetics." In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta, https://plato.stanford.edu/archives/sum2019/entries/environmental-aesthetics.

Coaffee, Jon, Paul O'Hare, and Marian Hawkesworth. 2009. "The Visibility of (In)security: The Aesthetics of Planning Urban Defences against Terrorism." *Security Dialogue* 40, no. 4–5: 489–511. DOI: 10.1177/0967010609343299.

Crippen, Matthew, and Vladan Klement. 2020. "Architectural Values, Political Affordances and Selective Permeability." *Open Philosophy* 3, no. 1: 462–477.

Den Tandt, Christophe. 1994. *The Urban Sublime in American Literary Naturalism*. Urbana IL: University of Illinois Press.

Den Tandt, Christophe. 2014. "Masses, Forces, and the Urban Sublime." In *The Cambridge Companion to the City in Literature*, edited by Kevin R. McNamara, 126–137. Cambridge: Cambridge University Press.

Dumbaugh, Eric, and J. L. Gattis. 2005. "Safe Streets, Livable Streets." *Journal of the American Planning Association* 71, no. 3: 283–300. https://doi.org/10.1080/01944360508976699.

Eames, Malcolm, Tim Dixon, Miriam Hunt, and Simon Lannon, eds. 2018. *Retrofitting Cities for Tomorrow's World*. Hoboken NJ: Wiley.

Epting, Shane. 2019. "Automated vehicles and transportation justice." *Philosophy & Technology* 32, no. 3: 389–403

Forsey, Jane. 2013. *The Aesthetics of Design*. Oxford: Oxford University Press.

Foster, Cheryl. 1992. "Aesthetic Disillusionment: Environment, Ethics, Art." *Environmental Values* 1, no. 3: 205–215.

Fox, Warwick. 2000. *Ethics and the Built Environment*. London and New York: Routledge.

Galloway, Alexander R. 2012. *The Interface Effect*. Cambridge: Polity.

Gehl, Jan. 1987. *Life Between Buildings: Using Public Space*. Translated by Jo Koch. New York: Van Nostrand Reinhold.

Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Godlovitch, Stan. 1994. "Icebreakers: Environmentalism and Natural Aesthetics." *Journal of Applied Philosophy* 11: 15–30.

Haapala, Arto. 2005. "On the Aesthetics of the Everyday: Familiarity, Strangeness, and the Meaning of Place." In *The Aesthetics of Everyday Life*, edited by Andrew Light and Jonathan M. Smith, 39–45. New York: Columbia University Press.

Haapala, Arto. 2017. "The Everyday, Building, and Architecture. Reflections on the Ethos and Beauty of Our Built Surroundings." In *Ethics in Architecture. Festschrift for Karsten Harries*, edited by Eduard Führ. Cloud-Cuckoo-Land, *International Journal of Architectural Theory* 22, no. 36: 169–182, http://cloud-cuckoo.net/fileadmin/issues_en/issue_36/article_haapala.pdf.

Hume, David. 1985 [1757]. "Of the Standard of Taste." In *Essays: Moral, Political and Literary*, edited by Eugene Miller. Indianapolis, IN: Liberty.

Ihde, Don. 1990. *Technology and the Lifeworld: From Garden to Earth*. Bloomington: Indiana University Press.

Jacobs, Jane. 1961. *The Death and Life of Great American Cities*. New York: Random House.

Jensen, Ole B. 2020. "Thinking with the Drone: Visual Lessons in Aerial and Volumetric Thinking." *Visual Studies*. doi: 10.1080/1472586X.2020.1840085.

Kant, Immanuel, 2000 [1790]. *Critique of the Power of Judgment* (*Kritik der Urteilskraft*). Translated by Paul Guyer and Eric Matthews. Cambridge: Cambridge University Press.

Kolhonen, Pasi. 2005. "Moving Pictures:- Advertising, Traffic and Cityscape." *Contemporary Aesthetics*, Special Vol. 1. https://contempaesthetics.org/newvolume/pages/article.php?articleID=351.

Leddy, Thomas. 2012. *The Extraordinary in the Ordinary: The Aesthetics of Everyday Life*. Peterborough: Broadview Press.

Lehtinen, Sanna. 2015. "Excursions into Everyday Spaces: Mapping Aesthetic Potentiality of Urban Environments through Preaesthetic Sensitivities." PhD diss. University of Helsinki.

Lehtinen, Sanna. 2020a. "Buildings as Objects of Care in the Urban Environment." In *Aesthetics in Dialogue: Applying Philosophy of Art in a Global World*, edited by Zoltan Somhegyi and Max Ryynänen, 223–236. Berlin: Peter Lang.

Lehtinen, Sanna 2020b. "Introduction to the Special Volume on Urban Aesthetics." *Contemporary Aesthetics*, Special Vol. 8. https://contempaesthetics.org/2020/07/16/editorial-introduction-to-the-special-volume-on-urban-aesthetics.

Lehtinen, Sanna, and Vesa Vihanninjoki. 2019. "Seeing New in the Familiar: Intensifying Aesthetic Engagement with the City through New Location-Based Technologies." *Behaviour & Information Technology* 39, no. 6: 648–655. https://doi.org/10.1080/0144929X.2019.1677776.

Lehtinen, Sanna, and Vesa Vihanninjoki. (forthcoming) 2021. "Aesthetic Perspectives on Urban Technologies: Conceptualizing and Evaluating the Technology-Driven Changes in the Urban Experience." In *Technology and the City: Towards a Philosophy of Urban Technologies*, edited by Michael Nagenborg, Margoth González Woge, Taylor Stone, and Pieter E. Vermaas. Dordrecht: Springer.

Light, Andrew, and Jonathan M. Smith, eds. 2005. *The Aesthetics of Everyday Life*. New York: Columbia University Press.

Lobo, Tea. 2020. "Urban Kinaesthetics." *Contemporary Aesthetics*, Special vol. 8. https://contempaesthetics.org/2020/07/16/urban-kinaesthetics.

Madsen, Peter, and Richard Plunz, eds. 2002. *The Urban Lifeworld: Formation, Perception, Representation*. London and New York: Routledge.

Manovich, Lev. 2001. *The Language of New Media*. Cambridge, MA: MIT Press.

Mattern, Shannon. 2018. "Maintenance and Care." *Places Journal* (November 2018). https://doi.org/10.22269/181120.

Mladenovic, Milos, Sanna Lehtinen, Emily Soh, and Karel Martens. 2019. "Emerging Urban Mobility Technologies through the Lens of Everyday Urban Aesthetics: Case of Self-Driving Vehicle." *Essays in Philosophy* 21, no. 2.

Mubi Brighenti, Andrea, ed. 2013. *Urban Interstices: The Aesthetics and the Politics of the In-Between*. London and New York: Routledge.

Naukkarinen, Ossi. 2005. "Aesthetics and Mobility: A Short Introduction into a Moving Field." *Contemporary Aesthetics*, Special vol. 1. https://contempaesthetics.org/newvolume/pages/journal.php?volume=13.

Nye, David. 1994. *American Technological Sublime*. Cambridge MA: MIT Press.

Nye, David. 2005. "The Sublime and the Skyline." In *The American Skyscraper: Cultural Histories*, edited by Roberta Moudry, 253–268. Cambridge: Cambridge University Press.

Parsons, Glenn, and Allen Carlson. 2008. *Functional Beauty*. Oxford: Oxford University Press.

Picon, Antoine. 2015. *Smart Cities: A Spatialised Intelligence*. Hoboken NJ: Wiley.

Pine, B. Joseph, and James H. Gilmore. 1999. *The Experience Economy*. Boston, MA: Harvard Business School Press.

Pyyry, Noora, and Sirpa Tani. 2016. "Young Peoples Play With Urban Public Space: Geographies of Hanging Out." In *Play and Recreation, Health and Wellbeing*, edited by Bethan Evans, John Horton, Tracey Skelton, 193–210. Singapore: Springer. https://doi.org/10.1007/978-981-4585-51-4_8.

Rice, Shelley. 1997. *Parisian Views*. Cambridge, MA: MIT Press.

Rosenberger, Robert. 2017. *Callous Objects: Design against the Homeless*. Minneapolis: University of Minnesota Press.

Rosenberger, Robert. 2019. "On Hostile Design: Theoretical and Empirical Prospects." *Urban Studies* 57, no. 4: 1–11.

Saito, Yuriko. 2007. *Everyday Aesthetics*. Oxford and New York: Oxford University Press.

Saito, Yuriko. 2020. "Aesthetics of Care." In *Aesthetics in Dialogue: Applying Philosophy of Art in a Global World*, edited by Zoltan Somhegyi and Max Ryynänen, 187–202. Berlin: Peter Lang.

Schummer, Joachim, Bruce MacLennan, and Nigel Taylor. 2009. "Aesthetic Values in Technology and Engineering Design," in *Philosophy of Technology and Engineering Sciences*, vol. 9, edited by Anthonie Meijers, 1031–1068. Amsterdam: Elsevier.

Shane, Grahame. 2002. "The Machine in the City: Phenomenology and Everyday Life in New York City." In *The Urban Lifeworld: Formation, Perception, Representation*, edited by Peter Madsen, and Richard Plunz, 227–236. London and New York: Routledge.

Shusterman, Richard, ed. 2019. *Bodies in the Streets: The Somaesthetics of City Life*. Leiden: Brill.

Simmel, Georg. 1969 [1903]. "The Metropolis and Mental life." In *Classic Essays on the Culture of Cities*, edited by Richard Sennett. Translated by H.H. Gerth, 47–60. Englewood Cliffs: Prentice Hall.

Stolnitz, Jerome. 1960. *Aesthetics and Philosophy of Art Criticism*. New York: Houghton Mifflin.

Stone, Taylor. 2017. "Light Pollution: A Case Study in Framing an Environmental Problem." *Ethics, Policy & Environment* 20, no. 3: 279–293.

Stone, Taylor. 2018. "Re-envisioning the Nocturnal Sublime: On the Ethics and Aesthetics of Nighttime Lighting." *Topoi*. https://doi.org/10.1007/s11245-018-9562-4.

Tainio, Matti. 2019. "Reconsidering Darkness." *Contemporary Aesthetics* 17. https://www.contempaesthetics.org/newvolume/pages/article.php?articleID=857.

Urry, John, and Jonas Larsen. 2011. *The Tourist Gaze 3.0*. London: Sage Publications.

Verbeek, Peter-Paul. 2005. *What Things Do. Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Pennsylvania State University Press.

Vihanninjoki, Vesa. 2020. "Atmospheric Affordances and the Sense of Urban Places." *Contemporary Aesthetics*, Special vol. 8. https://contempaesthetics.org/2020/07/16/atmospheric-affordances-and-the-sense-of-urban-places.

Vihanninjoki, Vesa, and Sanna Lehtinen. 2019. "Moving in the Metropolis: Smart City Solutions and the Urban Everyday Experience." In *Architecture and the Smart City*, edited by Sergio Figueiredo, Sukanya Krishnamurthy, and Torsten Schröder, 210–220. London and New York: Routledge.

Welsch, Wolfgang. 1991. *Ästhetisches Denken*. Stuttgart: Reclam.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109, no. 1: 121–136.

Wittingslow, Ryan Mitchell. (forthcoming) 2021. "Authenticity and the 'Authentic City.'" In *Technology and the City: Towards a Philosophy of Urban Technologies*, edited by Michael Nagenborg, Margoth González Woge, Taylor Stone, and Pieter E. Vermaas. Dordrecht: Springer.

PART VI

TECHNOLOGY, HEALTH AND THE ENVIRONMENT

CHAPTER 24

SCIENCE FICTION FUTURES AND (RE)VISIONS OF THE ANTHROPOCENE

JULIA D. GIBSON AND KYLE POWYS WHYTE

1. Introduction

The idea that global climate change—and the environmental injustices connected to it—signal or represent a new epoch of geological time has transfixed those in the sciences and the humanities. The Anthropocene, as some have proposed calling the moment we are now in or soon to enter, is a discourse that grapples with the ways in which human beings intentionally or inadvertently affect ecological systems on a global scale. A cluster of literatures, all deeply invested in diagnosing what went wrong and envisioning what can and ought to be expected of the future, has emerged alongside these ecological and ideological developments. As ecocritics have observed (e.g., Otto 2012; Gaard 2014; Rigby 2015), the genre of science fiction has established itself as a distinct facet of this conversation within and beyond the academy. Anthropocene futurities represent a particularly rich site of overlap connecting the philosophy of technology with philosophical and creative literatures on futurism, climate change, science fiction, and environmental justice.

Here, as elsewhere, technologies are often bound up with technological visions. Moreover, concepts of climate change or the Anthropocene are portrayals of technologies as technological systems. Such technologies include the physical infrastructure that drives certain anthropogenic forms of climate change, but also the associated human technological behaviors that are incentivized by culture, such as pollution. Andrew Feenberg claims, for example, that "technology is the medium of daily life . . . every major technical change reverberates at many levels, economic, political, religious, cultural" (1999, vii). In this sense of technology, visions tied to climate change or the Anthropocene present assumptions and beliefs about and aspirations for the future that affect multiple dimensions of society. Unfortunately, as with the Anthropocene discourse in general, much of the science/climate fiction that attends to environmental issues or crises has a rather singular vision of humanity, nature, apocalypse, and futurity that is incompatible with anti-colonial, anti-racist, and/or feminist approaches to environmental justice in the time of global climate change.

To a degree, philosophers of technology have played important roles in evaluating philosophically how futures are envisioned. For example, some philosophy of technology has examined how people's values shape their assumptions about future risks that they may be exposed to or how ethical processes should be established for people to gather and assess the weight of risks (e.g., Asveld et al. 2012; Floridi 2014). Philosophers have thought critically about how visions of the future motivate loss and change of certain traditional values and ethical commitments or create new forms of perception and cognition (e.g., Borgmann 1987; Ihde 1990). Sometimes research on future risks has been part of an analysis of environmental justice (Shrader-Frechette 2002), systems of power (Feenberg 1999), or sustainability (Thompson 2010). Importantly, work of this kind in the philosophy of technology has not fully examined the very nature of what it means to construct a technology future or vision of the future. Although risks, perils, and the foresight of harms are covered, there is little coverage of what such visions are or do in relation to technological systems like climate change. Often, technological visions are environmental visions, presupposing certain future states of ecosystems and the implications of those states of affairs for human existence and well-being. Risk perception often assumes certain beliefs, values, and knowledge about how the environment responds to pollution, for example; visions of technological transformation of traditional values are based on assumptions of how humans ought to relate morally and skillfully to non-human worlds.

One reason that more philosophical and expansive understandings of visioning and the future are absent from the philosophy of technology is the field's lack of diversity. Yet philosophers of technology can do more to ensure broader philosophical perspectives, especially ones that are not rooted primarily in a particular tradition of western philosophy (there are, of course, many traditions that might respectfully be called "western" in different ways). In the philosophy of technology, it is sometimes true that research on risk perception, visioning, and environmental justice relies on very particular thinkers or theoretical frameworks, such as Paul Thompson's focus on Thomas Jefferson's agrarian philosophy, Kristin Shrader-Frechette's avoidance of Indigenous and other non-western conceptions of justice, or Albert Borgmann's and Feenberg's being influenced by phenomenological and critical theory traditions emanating largely from Europe. Science fiction too has historically been less than friendly to women, persons of color, and others who would challenge its sense of realism and, relatedly, tends to rely on similarly narrow ideological frameworks.

The philosophies and narratives we are emphasizing here have different arguments and concepts pertaining to issues as diverse as knowledge and ethics; they also focus on different forms of oppression. For example, colonialism, as a form of oppression, is rarely taken up by philosophers of technology or in complex/intersectional ways by mainstream science fiction. Yet in Indigenous philosophy and science fiction alike, when the topic of oppression is discussed, colonialism is among the central topics.

And often colonialism is discussed intersectionally in relation to different forms of discrimination and violence, such as racism, patriarchy, and capitalism. Indigenous philosophical discussions about the future connect themes of technology, the environment, and colonialism, among others. The idea, for example, that climate change is an intensified form of colonialism certainly provides important insights—especially in contrast to how climate change is defined in other fora—about how some people think about the future in ways that draw out their experiences and point to gaps in the work of others. Colonialism, of course, is a technological system and has the features of technological systems and power that some philosophers of technology, especially Borgmann and Feenberg, have covered in their work. It is also true that while some philosophers of technology have taken up climate change, climate change is rarely (if at all) examined as a technological system itself in this literature. Strangely, even within the sub-genre of climate fiction or "cli-fi," this analysis is often lacking or seriously oversimplified.

In response to these limitations of the philosophy of technology and science fiction alike, this chapter explores the work of mainstream climate fiction in contrast with those of Indigenous, Afrofuturist, and/or feminist science fiction narratives. Rather than envisioning a monolithic cataclysm driven by technology or nature's whims, these alternative narratives situate environmental injustice and catastrophe within a complex web of intra- and inter-species politics. Their thoughtful world-building enables such stories to imaginatively mirror our own worlds such that the past, present, and future are transformed. In this way, Indigenous, Afrofuturist, and/or feminist science fiction narratives contribute ethical and political observations, theories, and visions crucial for the philosophy of technology.

2. Science Fiction on the Brink

Anthropocene discourse encourages us to think of humanity and the planet as being on the brink of a new epoch in which earth systems and thereby life will be radically altered. In geological terms, the proposed Anthropocene epoch is understood as a time in which the collective actions of humans began/begin to influence earth systems—including but not limited to climate—in marked, unprecedented ways. Though the precise start date and causes of the Anthropocene are continually up for debate, more recent theories link the proposed epoch to the onset of colonialism and global trade, particularly that of coal (Lewis and Maslin 2015). Since then, ever-expanding human economic activities and consumer lifestyles have become major co-drivers of ecological destabilization through their dependence on burning fossil fuels and certain kinds of land use; for example, deforestation. Scientists and environmental ethicists have tended to characterize Anthropocene futures in increasingly grim terms, most commonly by warning of or envisioning a world in which the very existence of certain ecosystems, plants, and animals is threatened by climate destabilization (Kolbert 2010; Thompson and

Bendik-Keymer 2012; Vaidyanathan 2014; Sandler 2014). Some conservationists argue that we will inevitably have to learn to live with these changes, make careful decisions about conservation priorities, and, in some cases, learn to let go of certain ecosystems and species (Kareiva and Marvier 2012). Yet others in the conservation community adamantly frame these losses, especially extinctions, as morally dreadful and, frequently, environmentally catastrophic (Vaidyanathan 2014; Cafaro and Primack 2014).

One of the authors (Whyte 2018) has written on some of the ways in which academics, journalists, artists, and writers alike have conjured apocalyptic and dystopian portrayals of perilous futures involving mass species extinctions, ecosystem degradation, and social upheaval. Much of recent science fiction—new or familiar—is adept at making readers feel as though the time, society, or the planet in/on which they live is balanced on the edge of a knife. The drama of such stories revolves around characters working to prevent, manage, navigate, or survive the tipping point (e.g., Garrard 2012; Otto 2012; Morton 2013; Gaard 2014; Anson 2017; Whyte 2018). The trope of the tipping point is especially prevalent in the sub-genre of climate fiction, whose narratives tackle the threat or reality of global climate change (more or less) head-on. Kim Stanley Robinson's (2004–2007) *Science in the Capital* trilogy, for example, revolves around the lives and efforts of scientists and politicians in early 21st-century Washington, D.C. to warn of, stave off, and eventually mitigate or adapt to anthropogenic climate change by advancing (primarily) technological solutions. Due to their specificity and particular style of realism, cli-fi narratives like Robinson's are particularly adept at cultivating the impression that the tipping point is *right now* and that very soon everything could change or fall apart.

In the context of most mainstream sci-fi and, in particular, cli-fi, the brink is a threshold that should not be crossed at all costs. Beyond this tipping point lies apocalypse, the end of the world. The apocalypse of such narratives manifests as societal collapse, ecological collapse, or some combination of the two (Otto 2012). Causes for collapse vary, and the triggers for tipping points and/or apocalypse are imaginatively depicted in several key—and often overlapping—ways:

Extraction-driven ecological devastation; for example, *The Lorax* (Seuss 1972), *Avatar* (Cameron 2009), *Fern Gully* (Kroyer 1992), and *Dune* (Herbert 1965)

Natural catastrophes or acts of god; for example, *Armageddon* (Bay 1998), *Noah* (Aronofsky 2014), and *Interstellar* (Nolan 2014)

Pollution; for example, *Once Upon a Forest* (Grosvenor 1993)

Nuclear fallout or winter; for example, *Doctor Strangelove* (Kubrick 1964), *Z for Zachariah* (O'Brien 1974), and *The 100* (Morgan 2013–2016)

Weakness or evil of humans; for example, *Noah*, *Doctor Strangelove*, and *The Bone Clocks* (Mitchell 2014)

Technology run amok; for example, *The Day After Tomorrow* (Emmerich 2004) and *The Carbon Diaries: 2015* (Lloyd 2009)

The apocalypses depicted in mainstream science/climate fiction are a mix of "tragic" (i.e., apocalypse is inevitable and redemption from human guilt or evil is to be found in sacrifice) and "comic" (i.e., apocalypse is avoidable and redemption from human error is to be found through recognition), with an emphasis on the latter (Garrard 2012). Either way, however, apocalypse lies in the future, just over the horizon.

Ecocritics, at times reluctantly, recognize the power of the apocalyptic trope—presumably in its comic form—to awaken readers to their close proximity to the brink and thereby inspire action (Garrard 2012; Otto 2012; Schatz 2012). Framing environmental quandaries and losses through the lens of apocalypse imparts a sense of urgency that can be distinctly motivating. The sense readers get from the bulk of these narratives is that humanity can pull itself and, thereby, the world back from the brink of apocalypse—or, at least, avoid the worst of it—if only we put our minds to it and try hard enough. It's not too late, not yet. Likewise, ecocritics maintain that many works of science/climate fiction articulate innovative solutions of their own for averting the worst of apocalypse and surviving the rest. Whether or not they locate the causes of apocalypse with technology, much of this genre skews strongly toward technologically oriented solutions and strategies. When it comes to averting or forestalling ecological and societal collapse, science/climate fiction tends to advocate either for embracing—for example, *Star Trek* (Roddenberry 1966–1969, 1987–1994) and *Armageddon* (Bay 1998)—or abandoning technology; for example, *The Queen of the Tearling* series (Johansen 2014–2016). In the science fiction disaster film *Armageddon*, for example, a team of deep-sea oil drillers detonates a thermonuclear bomb in an asteroid hurtling toward earth in order to avert a cataclysmic extinction event. However dubious the science and overblown the narrative, the film aptly portrays the belief that—in the hands of good men—technology can surely overcome any threat to humanity.

Other science/climate fiction narratives, such as the *Science in the Capital* trilogy (Robinson 2004–2007), explore what surviving the apocalypse could look like. Much like stories in which apocalypse is averted, survival narratives often foreground exaltations or critiques of technology. Robinson's trilogy is of the former type and strongly advocates for science and technology to take charge in the face of apocalypse. Yet despite the fact that climate change is far from averted in this series and ecological changes abound (e.g., the Gulf Stream has stalled out), societal changes are distinctly muted. Nations, gender roles, institutions (governmental agencies and NGOs alike), racial categories, international governing bodies, economic systems, etc. all remain largely unchanged. In this way, Robinson's trilogy aptly demonstrates the tendency of mainstream science/climate fiction to "end" the world while managing to keep it recognizable to environmentally privileged readers in the global North. Such narratives travel past the brink and yet somehow fail to fall into it.

Through their exploration of what surviving environmental apocalypse could look like, science/climate fiction narratives also weigh in on or even theorize about what parts of the world are worth salvaging and what and who are not. As we have seen, modern technology may or may not be worth carrying forward into the future, but it is not the only aspect of the pre-apocalyptic world to be considered and ultimately rejected or accepted onto the ark. Contemporary gender roles, capitalist economic arrangements, settler-coloniality, modern racial hierarchies, among others, are all recurrent ticketed passengers on the ark or manage somehow to hitch a ride. Even when mainstream narratives frame apocalypse as more or less inevitable and devastating, conflict tends to revolve around the fight for certain ideological fixtures of western democratic societies. Preserving the last shreds of human freedom and humanity itself, for instance, are crucial priorities in the futures envisioned by tragic science/climate fiction; for example, *Planet of the Apes* (Schaffner 1968) and *Mad Max: Fury Road* (Miller 2015). Whether by building an ark or by fighting to keep afloat precious flotsam in the storm, these texts make normative and political claims regarding the types of futures that should be desired and aimed for (Otto 2012; Tuck and Yang 2012; Gaard 2014).

Although science/climate fiction does not always explicitly reference the Anthropocene by name, these narratives feed off of and grapple with the same anxieties regarding global climate change and other anthropogenic environmental harms and injustices that preoccupy the nonfictional literatures of the Anthropocene discourse. Moreover, these stories and the broader discourse are placed into conversation by ecocritics (e.g., Otto 2012; Rigby 2015), environmental ethicists (e.g., Whyte 2018), and those in the environmental humanities (e.g., Anson 2017). Perhaps the most important overlap for our purposes, however, lies with the value assumptions that mainstream science/climate fiction shares with the broader discourse. In the next section, we explore how the values that shape and motivate the futures depicted by these narratives serve as an expression of Anthropocene futurity.

3. A Singular Vision

When conceived and executed well, science fiction can gesture toward or suggest ways to grapple with environmental injustice, technological challenges, and climate change. Such guidance is likely to go awry, however, when narratives rely upon problematic value assumptions and incomplete or inaccurate descriptions of the politics and technologies behind the anthropogenic phenomena in question. Ursula Le Guin (1976b) reminds us that science fiction is a descriptive—rather than predictive—endeavor but strongly cautions (2004, 218–219) against "timid and reactionary" science fiction fantasy storytelling. She writes, "The imaginative fiction I admire presents alternatives to the status quo which not only question the ubiquity and necessity of extant institutions, but enlarge the field of social possibility and moral understanding" (Le Guin 2004, 219–220). Although mainstream apocalyptic science/climate fiction may indeed succeed in waking some readers up to the urgency of potential and ongoing environmental and social ills, they tend to offer a problematically totalizing, reductive description and vision. As with the Anthropocene discourse at large, it is easy to walk away from these narratives with the impression that there is but one humanity, one nature, one apocalypse, one history, and one future. This section considers each of these normative and/or political assumptions in turn while also attending to the ways in which they influence each other.

Mainstream science/climate fiction has an unfortunate tendency to gloss over or fail to attend to considerations of gender, race, sexuality, disability, and, in particular, coloniality. Moreover, when such narratives do explore social dimensions (most commonly, class) in more depth, intersectional dynamics are still largely neglected (Gaard 2014). As a result, environmental threats and apocalypse get framed as problems that all of humanity (must) face together. The existence and exacerbation of preexisting vulnerabilities and the disproportionate impact of environmental crises such as global climate change are often lost (Cuomo 1998; Morton 2013; Rigby 2015). Likewise, culpability for "anthropogenic" crises is generally portrayed somewhat simplistically and without consideration for historical and ongoing violence and oppression. Just as humans are all in the same boat when it comes to confronting environmental calamity together, "we" are also all to blame as a species for mucking up the environment, climate, biosphere, etc. (Whyte 2017). The "we" that is given face, voice, and agency in science/climate fiction, however, tends to be white, male, environmentally privileged, and technologically "advanced" (Gaard 2014). Thus, these narratives frame humanity as a whole as at odds with nature (writ large) in terms of both vulnerability and culpability, while often only considering a narrow slice of human experiences and identities.

Nature, for its part, gets defined in opposition to or in contrast with humanity and technology (Cuomo 1998; Haraway 2008; Gaard 2014). That being said, there are many different instantiations of the nature versus humans/technology trope that crop up throughout science/climate fiction, Anthropocene literature, and beyond. Some narratives reduce nature to a set of natural resources that humans (via technology) must make sure to use wisely. Thusly instrumentalized and de-animated, nature recedes into the background, a cluster of material resources that need not take up any further human attention if their supply is not threatened. When nature is not more or less backgrounded, it tends to take on the narrative role of the victim or villain against which humans/technology must prevail. As a victimized object/entity—often femininely gendered—nature is romanticized and reduced to a passive system of species and habitats that humans have the unique responsibility to save. In this framing, the loss of nature is tragic given its innocence and lack of complicity in its own destruction. Often these renditions of nature suggest humans derive spiritual sustenance from features of nature, one common example being a giant or sentient tree; for example, *Pocahontas* (Gabriel and Goldberg 1995) and *Avatar* (Cameron 2009). As a villain, nature can be reduced to a dormant power whose fury and violence are unleashed when humans abuse or neglect it. Species, elements, or systems may all of a sudden overwhelm humans when they fail in their responsibilities. Such fictional accounts of nature often suggest a human ambivalence toward nature, in the sense that nature cannot ultimately be trusted and nature has no accountability to human life. Of course, such framings of nature can be mixed together. Sometimes the passive, romanticized nature is pushed too far and transforms into an aggressive villain. Other times, the various notions stay in their own lanes.

Ultimately, such conceptualizations reinforce the idea that nature (singular) interacts with a humanity (also singular) and the technology at "our" disposal. Most fictional accounts, for example, portray humans as responding to a monolithic nature, whether instrumentalized or victimized, romanticized or villainous. Some science/climate fiction accounts complicate this picture by orienting the narrative around conflict that has arisen between different groups of people who view nature and technology differently, such as conflicts between those who value nature as a pure instrument and those who invest spiritual value in nature. Unfortunately, these narratives generally resolve with all sides coming to appreciate—as humans united—the same view of nature and/or the proper uses of technology, as in stories involving purely instrumental valuers of nature realizing that the non-human world is indeed sacred. Other times, environmental apocalypse can only be averted when the proper balance between nature and technology is arrived at either by consensus or, more commonly, heroic action/force (Gaard 2014). Rarely is the "problem" of technology framed in terms other than the root cause of or sole hope to avoid environmental catastrophe.

When narratives portray a single humanity and technology interacting with a single nature, environmental catastrophe can likewise be conceptualized as singular and, all too often, monolithic in nature. Apocalypse, then, is the destruction of a single world, which here refers to the coupling of human and natural systems. The presumption of a singular shared world supports the idea that both science/climate fiction apocalypses and the nonfictional environmental catastrophes they "describe" are unprecedented in human and, as anthropogenic phenomena, planetary history. Add to that the failure to grapple with varying degrees and kinds of vulnerability and culpability across space, time, embodiment, and identity and the causal mechanisms of apocalypse tend to get obscured. Oversimplified or confined to ill-fitting allegories, the fictional causes of environmental destruction and destabilization do not typically map well onto reality. Likewise, such apocalypses are temporally distorted. Science/climate fiction apocalypses tend to happen for everyone all at once, with the time leading up to the tipping point being relatively stable, however imperfect. Although inequalities may exist in pre-apocalyptic science/climate fiction worlds, these narratives typically give the impression that what looms on the horizon (or has just hit home) is like nothing anyone has seen before. Apocalypse is new and it is now.

Coupled with the oversimplification and misrepresentation of apocalypse's agents, causes, and victims, this temporal collapse frequently results in monolithic science fiction futurities. (Post)apocalyptic narratives are generally designed to make the reader or viewer feel anxious or unsettled about the past and present through the guise of the future. That being said, while science/climate fiction may not be predictive, the genre can and does offer warnings about what is likely to happen in the future if the status quo persists. Even then, however, such warnings are perhaps better read as critiques of the present—via counterfactual thought experiment (Whyte 2018)—than commentary on the future itself. One way in which science and, in particular, climate fiction can weigh in on the future is *prescriptively*. When it comes to the world's future the commentary articulated by and through science fiction concerns not what *will* be but what *could* be. Such stories ask—and sometimes answer—questions regarding the kind of worlds we ought to be building. But with the "we" of humanity and the world it occupies framed so singularly or monolithically, the futures imagined in science/climate fiction are similarly warped. When humanity wears a problematically narrow range of faces, so too do the denizens of future worlds. The most generous interpretation of such futurities would be to say that they are incomplete. A more critical analysis reveals that a byproduct—or, perhaps, implicit goal—of many dystopian and/or (post) apocalyptic narratives is the rescuing of white, settler, environmentally privileged, etc. futurities (Tuck and Yang 2012: Gaard 2014). Although these communities are disproportionately responsible—in the real world if not in fiction—for environmental destabilization and injustice, when writing themselves into the future few stop to question whether they should be there and, if so, in what forms. Mainstream science/climate fiction thereby refuses to contemplate a world in which the communities, values, technologies, and lifeways responsible for the cataclysm do not survive unchanged or at all.

These monolithic conceptualizations of humanity, technology, nature, apocalypse, and future are at odds with the complex politics responsible for and expressed by the phenomena known collectively as the Anthropocene. Crucially, however, much of what is troubling about mainstream science/climate fiction is precisely what many have already identified as being problematic about the Anthropocene discourse and the concept itself. As critical Anthropocene scholars have argued, a single Anthropos does not exist (Cuomo 2011, Gaard 2014; Haraway 2015). Likewise, a single nature with which to contrast it does not exist (Plumwood 1993; Lepori 2015; Vogel 2015). Furthermore, vulnerability to and responsibility for global environmental injustices vary widely from community to community, human or otherwise (Cuomo 2011; Lepori 2015). In fact, disproportionate vulnerabilities and responsibilities are a defining feature of climate change, thus working against the idea of there being *an* Anthropocene, as does the reality that environmental apocalypse has already happened or is currently happening—for the first time or all over again—for many, in particular Indigenous/colonized communities and the descendants of enslaved peoples.

Heather Davis and Zoe Todd see an insidious irony in the different ways Indigenous and non-Indigenous persons approach the Anthropocene and climate crisis. They describe colonialism as a seismic shockwave that "kept rolling like a slinky [as it worked] to compact and speed up time, laying waste to legal orders, languages, and place-stories in quick succession. The fleshy, violent loss of 50 million Indigenous peoples in the Americas is something we read as a 'quickening' of space-time" in a seismic sense (Davis and Todd 2017, 771–772). Davis and Todd then point out that "the Anthropocene or at least all of the anxiety produced around these realities for those in Euro-Western contexts—is really the arrival of the reverberations of that seismic shockwave into the nations who introduced colonial, capitalist processes across the globe in the first half-millennium in the first place" (Davis and Todd 2017, 774). Although frequently framed as an epoch of unprecedented human flourishing and technological advancement (Rockstrom et al. 2009), the (tail "end" of the) Holocene was an especially brutal time for much of humanity. The entire endeavor of carving up geologic time into planetary epoch, eras, etc. has also been critiqued as a deeply colonial project (Cuomo 2014; Mitchell 2015; Davis and Todd 2017). All in all, it is perhaps not surprising that fictional narratives attempting to describe and respond to the Anthropocene end up reproducing similar frameworks and logics with regard to time, place, technology, and life.

4. Visionary Alternatives

Fortunately, there are numerous science fiction writers and artists who do not couch their observations, critiques, and recommendations for redressing environmental apocalypse and injustice in terms of the Anthropocene. Instead, their narratives articulate values, descriptions, and futurities that stand in stark contrast to those commonly employed within mainstream science fiction and Anthropocene discourse alike. The stories discussed in this section are what Adrienne Maree Brown and Walidah Imarisha[1] (2015) would describe as "visionary fiction," that is, "a term . . . developed to distinguish science fiction that has relevance toward building new, freer worlds from the mainstream strain of science fiction, which most often reinforces dominant narratives of power" (4). Visionary fiction encompasses stories within science fiction, fantasy, speculative fiction, magic realism, etc. whose purpose is social change and transformation. The elements of visionary fiction include exploration of current social issues; consciousness of identity and intersections thereof; the centering of those who have been marginalized; an awareness of power inequalities; demonstration of change from the bottom up achieved collectively; and realism that is hard but hopeful (279).

In recognition of the fact that critiques of Anthropocene discourse tend to be mobilized along lines of gender, race, and/or coloniality, narratives in this section are organized under these headings. This is not to suggest that these are the only salient dimensions of identity and power when it comes to environmental injustice. Likewise, the intention is not to imply that these "axes" of oppression do not intersect. They surely do. Each subsection here is intended to build upon the last, and our effort to achieve focus should not obscure the reality of interconnection and intersection. Nevertheless, many science fiction narratives—even visionary ones—tend to focus on certain intersections and power dynamics more than others. This can be, although certainly not always, done well; that is, in ways that do not serve to erase or perpetuate violence along backgrounded or secondary axes. Even when well implemented, however, the results often highlight what Tuck and Yang (2012) refer to as the incommensurability of different anti-colonial and social justice projects. The division of texts in this section is also intended to help make that incommensurability—and the resulting implications for solidarity—more visible.

5. Visions of Gendered Environmental Injustice, Resistance, and Liberation

Much praised and analyzed, Margaret Atwood's (2003–2013) *MaddAddam* trilogy is perhaps the best-known example of feminist climate fiction (e.g., Ullrich 2015; Traub 2018). Atwood's trilogy details the post-apocalyptic struggles of the last human(s) on earth following a plague engineered by the world's best scientist(s) to wipe the species from the planet. Our protagonists' and anti-hero's worlds are not small, however. Through frequent flashbacks we learn of the time before "the flood" in all its glorious, heart-rending detail. The pre-flood world is both deeply disturbing and utterly recognizable. The rampant abuses of capitalism and technology, gendered and racialized violence and inequality, and increasingly destabilized climate all seem like the next logical incarnation of the environmental, gender, economic, and racial injustice that abound in today's world. The plague may have decimated humanity, but Atwood makes clear that its development and implementation are but one strand in the apocalyptic web. Moreover, the plague is the least of the characters' concerns when it comes to surviving amidst and upon the detritus of a world torn asunder and extremely reluctant to die.

One aspect of these novels that has made them successful and useful for theorizing climate justice is Atwood's skill at depicting characters "dancing with" and adapting to disaster as it unfolds over time (Rigby 2015). As in the pre-flood world (although to varying degrees and in various ways), nothing about their survival is assured. To the very end, these novels leave the fate of the protagonists' efforts to (re)establish community uncomfortably uncertain. Without being saccharine, however, the trilogy vividly conveys how worthwhile and beautiful the work of transformation can be in spite of this. Unfortunately, Atwood's trilogy also falls prey to some of the same problems that crop up in mainstream climate fiction. For one, there is very little (de)colonial awareness; Indigenous people simply are not present in either the pre-flood or post-flood worlds. Not only does this make Atwood's narratives descriptively inadequate, but also the futurities represented therein are thereby suspect. What does it mean for a ragtag bunch of former sex workers, anarchists, and hackers to survive alongside a new genetically engineered sapiens species and human-pig hybrids on the eastern seaboard of North America when (apparently) the Indigenous inhabitants of this place did not survive? Many readers of the *MaddAddam* trilogy will not even think to ask this question; the narratives do nothing to prompt it.

By contrast, Ursula Le Guin's (1976a) *The Word for World Is Forest* certainly does not neglect the racial and colonial dimensions of environmental (in)justice. Despite being published several decades before the Anthropocene was formally conceptualized, Le Guin's novella thoughtfully explores how colonial logics and epistemologies of ignorance make technology-intensive, extractive capitalism seem like the only viable option. Set on the alien planet of Athshe, the narrative centers around three characters: Captain Davidson, the military commander of a human logging operation; Raj Lyubov, the mission anthropologist, also human; and Selver, a native Athsean and formerly a slave in a logging camp, whose wife is raped and murdered by Davidson. While the native Athseans are humanoid, they are considerably smaller and furrier than humans. These qualities—and their seeming laziness and lack of ability or inclination to make use of the forest—lead the human colonists to believe they are sub-persons who can be justifiably enslaved. Having come to know Selver in the camp, Lyubov begins to doubt this assessment at the same time that Selver organizes an insurrection against the colonists. Despite Lyubov's entreaties, Davidson refuses to halt logging and the Athseans are forced to resort to violent tactics not previously practiced in their society. Following their victory, Selver remarks of the future he helped create, "Sometimes a god comes ... He brings a new way to do a thing, or a new thing to be done. A new kind of singing, or a new kind of death. He brings this across the bridge between the dream-time and the world-time. When he has done this, it is done. You cannot take things that exist in the world and try to drive them back into the dream, to hold them inside the dream with walls and pretenses. That is insanity. What is, is. There is no use pretending, now, that we do not know how to kill one another" (Le Guin 1976a, 188–189).[2]

Although critiqued for not being as complex as some of Le Guin's other science fiction works, *The Word for World is Forest* offers an unflinching examination of the violence (technological and otherwise) inherent in colonialism, anthropocentrism, racism, sexism, and capitalism and of the linkages between them. Much of this is illustrated through Davidson's characterization and inner monologue. There is no doubting that this antagonist's views about race, gender, and nonhumans are responsible for his inability to see Athshe—and its humanoid inhabitants—as anything other than resources for the taking. Although Davidson is depicted as unquestionably loathsome, he is not a one-dimensional character. His villainy may not be remotely ambiguous, but the reader is provided with extensive detail regarding how and why Davidson is the way he is. Through Lyubov's practice of anthropology, Le Guin is also careful to explore how science and technology are implicated in colonialism, as well as potential resources and sites within science for doing anti-colonial work. Thus, beyond the descriptive adequacy of the text, it articulates theories for understanding and resisting environmental injustice in the world beyond the page. The aspect of the novella that may be troubling for some readers is the fate of the Athseans. As the coordinator of the rebellion and god of war, Selver feels certain that even after the Athseans regain their forest, life within it will never be the same for having known such deliberate violence. In addition to the questionable decision on the part of a non-Indigenous author to circumscribe Indigenous futurity so definitively, one might wonder what makes the violence of liberatory insurrection temporally distinct (i.e., a "there's no going back" affair) from that of rape and slavery under colonialism.

The line between colonist and colonizer is somewhat murkier, although no less important, in Kameron Hurley's (2017) *The Stars Are Legion*. A bizarrely brilliant space opera, Hurley's novel takes place in the outer reaches of a fictional star system populated by living "world-ships"—collectively known as the Legion—and their all-female inhabitants. The accelerating decay of these living vessels/planets has led to perpetual conflict among the surface-dwelling humanoid "rulers" (and their armies) of various planetary clusters. Star-crossed lovers Zan and Jayd aim to put things to rights by obtaining access to a world-ship rumored to possess the power to regenerate itself and other worlds. The crucial problem—and the driving narrative force of the novel—is that Zan has recently been resurrected from the dead with (intentionally) little memory of her previous life/lives. After Jayd is married off to broker peace, Zan finds herself driven to the (living) core of her world-ship and undertakes a perilous journey back to the surface. Along the way she encounters numerous allies and foes among the societies that call the various subterranean levels of the world-ship their home, many of which doubt the very existence of the surface Zan seeks. As she climbs, Zan gradually regains piecemeal memories that indicate this isn't the first time she's encountered the lower levels of the world-ship, causing her to doubt Jayd, their mission, and who she understands herself to be.

Hurley's choice to populate the Legion exclusively with women could easily have backfired spectacularly. Instead of a simplistic utopic vision, however, *The Stars Are Legion* offers a (literally) multilayered apocalyptic landscape that "imaginatively mirrors" the politics of climate change all while retaining its gendered realities despite the total absence of men (Little 2007). For example, all women of the Legion have wombs but each give birth to different sorts of entities (e.g., organic hardware, monstrous creatures, food) that/who are of more or less use to the world-ships and the people and societies who call them home. In addition to framing the womb as a site of technology, the novel encourages readers to contemplate (re)production in a context in which birth is not the purview of an "inferior" gender. But as Hurley's narrative suggests, on/in worlds where colonial logics produce violent, unsustainable forms of life and death, the politics, ecology, and technologies of birth are no less disturbing. And yet hope remains. Indeed, the novel produces bold feminist futurities that revolve around the nexus of memory, birth, death, and loss. For both Zan and others, memory is dangerous, emotionally devastating, and necessary for building a better world. At the end of her journey Zan narrates, "We are two women standing at the edge of the Legion, our armies dead, our people broken, with a history between us that I no longer want filled in any further. Instead, in my mind I construct a future . . . It's a potential future for us, as real as the potential of the child I sacrificed to get here, as real as the dreams of the people who helped to get me this far" (Hurley 2017, 380). The future can neither dwell in nor forget the past, no matter how much it might want to.

6. Visions of Racialized Environmental Injustice, Resistance, and Liberation

Afrofuturist and climate fiction classics, Octavia Butler's (1993, 1998) *Earthseed* duo—*Parable of the Sower* and *Parable of the Talents*—tell of the life of Lauren Olamina, founder of the Earthseed religion. Having been raised in a gated community turned semi-commune, Lauren is spared from the worst of southern California in the 2020s in the wake of climate, political, and economic destabilization. Born with "hyperempathy," Lauren is able to share the emotions and sensations of others in close proximity, thus making outings beyond the walls of her community into the city extremely unpleasant and often painful. In other stories hyperempathy might be counted as a blessing, but for a young Black woman in a dystopic society rife with suffering, Butler is clear to frame it as a liability. As she grows up, Lauren is unsatisfied with her father's Baptist teachings and, instead, begins to imagine a faith organized around the principle of change and the idea that humans are destined to leave the planet. When outsiders attack and destroy her home, Lauren flees north to begin again, drawing followers with her talk of Earthseed along. The group founds the community of Acorn and lives happily for several years. In the conclusion of the series, however, Acorn is occupied by Christian fundamentalists—emboldened and empowered under a xenophobic zealot in the White House—who separate the children (including Lauren's) from their parents and place them in Christian homes. The residents of Acorn eventually rise up and escape their captors, but Lauren is unable to reunite, if not reconcile, with her daughter until much later. By the end of Lauren's life, Earthseed is flourishing and human settlers are traveling into space.

With Earthseed, Butler highlights both the destructiveness and necessity of change. As she explains in a 1999 interview, "Lauren Olamina says that since change is the one inescapable truth, change is the basic clay of our lives. In order to live constructive lives, we must learn to shape change when we can and yield to it when we must. Either way, we must learn and teach, adapt and grow" (Butler 1993, 336). When Lauren's first community was destroyed she founded another, bringing with her what she valued about the old and discarding the rest. Indeed, the theme of community is perhaps just as important to the Earthseed series and religion as change. These narratives highlight the material, emotional, and spiritual necessity of community, as well as the ways in which oppressive power-structures seek to undermine it. Between her fledgling Earthseed faith and the value she places on building and maintaining community, Lauren quickly learns to be an activist, embodying what it means to be a prophet for apocalyptic times through ideological and tangible ways.

Tan-Tan, the protagonist in Nalo Hopkinson's (2000) *Midnight Robber*, is also a harbinger of change. The novel takes place on the planet of Toussaint, an alien world settled by the survivors of white imperialism and colonialism who left Earth to start anew. This world, however, was not empty upon their arrival, but by the time Tan-Tan is born, all of the remaining Indigenous inhabitants of Toussaint seem to have been relegated to a mirror dimension—New Half-Way Tree. The daughter of a wealthy and powerful man who commits murder and is sentenced to exile, young Tan-Tan finds herself on New Half-Way Tree when her father (illicitly) takes her with him. Upon arrival, Tan-Tan meets the douen Chichibud—a native of the place the humans call New Half-Way Tree—who guides her to a human settlement and, years later, takes a pregnant Tan-Tan to live with his family after she kills her abusive father. In the village, Tan-Tan is trusted to learn and keep the secrets of the douen, who have successfully managed to hide many aspects of their existence from the (unwilling) colonists. But willing or not, most of the humans consider the douens an inferior species and pose an increasing threat to them and their way of life. Suffering from trauma and the foolishness of adolescence, Tan-Tan convinces Chichibud's young daughter, Abitefa, to help her implement vigilante justice throughout the human settlements as the Robber Queen, eventually leading enemies back to douen. As a result, the village is tragically forced to relocate and Tan-Tan and Abitefa are made to live on their own. In the end, the birth of her child (and sibling) forces Tan-Tan to confront her external and internal demons. She rejoins human society but retains the mantle of Robber Queen, working always to build the life that she and all the inhabitants of New Half-Way Tree—douen, human, or otherwise—deserve.

Through *Midnight Robber*, Hopkinson constructs a fascinating context for contemplating the relationships between settler and Indigenous persons and communities, especially those involving unwilling colonizers. When pushed through the dimensional veil, the new human inhabitants of New Half-Way Tree bring with them all sorts of hitchhikers. Some manifest as "invasive" species (e.g., grains, fruits, livestock). Interestingly, however, the douen are not engaged in efforts to eradicate these new lifeforms, even though they can be disruptive. Instead, they work to incorporate them into native ecosystems and develop relationships with them such that they can leverage more power among the humans. Other tag-alongs are not so easy to work with. So pervasive are gendered, raced manifestations of colonial violence that female douens—large birdlike creatures very different in form from the more humanoid males—do not reveal themselves to humans as either members of the same species or capable of speech. Similarly, the douens do not share much of their knowledge of the forest flora and fauna with the humans out of concern for the way these exiles have cultivated extractive relationships with their environs. Only Tan-Tan, who comes to New Half-Way Tree as a small child, questions the corrosive social norms of the penal colony—including the subhuman categorization of the douens—enough to learn from the Indigenous inhabitants and attempt to foster new ways of life among the humans. And through Tan-Tan, Hopkinson provides a critique of the tensions between formerly enslaved persons, communities, and native peoples, as well as a possible roadmap for navigating them moving forward.

In contrast with Atwood's series, N. K. Jemisin's (2015–2017) *Broken Earth* trilogy is a science fiction tour de force that heartbreakingly highlights the intersections of gendered, racialized, colonial, heteronormative, and environmental violence and injustice through the lens of Afrofuturism. These novels take place in a world of tectonic upheaval literally held (mostly) together by an enslaved class of humans with the ability to work magic on rock and earth. The efforts of these mages or "orogenes," however, is not enough to hold back massive geologic ruptures that the "evil earth" manages to unleash every few hundred years that trigger cataclysmic climate changes or "fifth seasons." As a result, the dominant society has been organized around making oneself and one's community as fit as possible in preparation. The events of the trilogy begin with the deliberate triggering of an unprecedentedly devastating fifth season and a father's murder of his young son, who is discovered to be an orogene. The plot of the trilogy follows the boy's mother, Essun, in search of her daughter, Nassun, who has been abducted by her father in the wake of the murder. Unfolding across a vast supercontinent and various decades and millennia, Essun's and Nassun's stories force readers to confront the repeated world endings experienced by enslaved and marginalized persons, as well as the question of whether those whose worlds have ended repeatedly have any obligation, given the choice, to keep the larger world from burning. This choice is put before several orogenes throughout the novels, ultimately culminating in Nassun's decision to allow the scattered fragments of humanity to remake the world together.

The Fifth Season (Jemisin 2015) opens with the passage, "Let's start with the end of the world, why don't we? Get it over with and move on to more interesting things . . . But this is the way the world ends. This is the way the world ends. This is the way the world ends. For the last time." And even though by the end of the trilogy this passage takes on a great deal of nuance, Jemisin never abandons her critique of frameworks—like the Anthropocene—that are unable to accommodate the complex temporality, spaciality, and subjectivity of apocalypse. This idea is echoed in Kathryn Yusoff's (2018) work entitled *A Billion Black Anthropocenes or None*, which builds off of Jemisin's narratives and Black feminism more generally. These novels situate climate change as one among many sorts of apocalypse to unfold within and from complex assemblages of oppressive power structures. The world has ended just as surely when a young Essun takes the life of her own child rather than see him an enslaved orogene like she was, as it ends years later when the child's father, Alabaster, tears a continent asunder. And so when Essun discovers that the (sentient) Earth is just another parent whose child (the moon) has been ripped away from them, she fights tooth and nail for a solution that will see them both reunited with their offspring for a future in which both can flourish.

7. Visions of (De)colonial Environmental Injustice, Resistance, and Liberation

Indigenous peoples have already endured harmful and rapid environmental transformations due to colonialism and other forms of domination.[3] As Davis and Todd (2017) articulate so clearly, these environmental transformations—"the fleshy violent [losses]"—seem actually a lot like what many other people in the world fear will happen with climate destabilization when these same people portray apocalyptic and dystopian science fiction futures. Whyte cites Lee Sprague, who says that we already inhabit what our ancestors would have understood as a dystopian future (Sprague 2017; Whyte 2017). Larry Gross writes that "Native Americans have seen the end of their respective worlds . . . Indians survived the apocalypse" (2014, 33). Sprague's and Gross's framing of today's times come out in Indigenous science fiction expression. In her short story anthology *Walking the Clouds*, Grace Dillon (2012) interprets Indigenous futurisms in literature and the arts as expressing how Indigenous peoples are currently living in a "post-Native Apocalypse" (Dillon 2012, 10). Building on Dillon's research, Conrad Scott's recent study discusses how "Indigenous literature, following the culturally destructive process of colonial European advancement and absorption of what are now called the Americas, tends to narrate a sense of ongoing crisis rather than an upcoming one" (Scott 2016, 77).

Cutcha Risling Baldy describes Indigenous histories and experiences of colonialism as suffering through the television zombie series *The Walking Dead* (Risling Baldy 2014). It is not hard to see why historic and contemporary persons and institutions who participate in settler colonialism are not different from a zombie apocalypse. Like in dystopian science fiction, our ancestors would have seen us living in a situation in which the conditions of our individual and collective agency are almost entirely curtailed. But our ancestors and future generations are rooting for us to find those secret sources of agency that will allow us to empower protagonists that can help us survive the dystopia or post-apocalypse. And there is quite a bit of creativity involved in figuring out who the protagonists will be. The literature on Indigenous science fiction discusses the range of protagonists that Indigenous authors introduce in their narratives, from non-humans to spirits to women to youth (Dillon 2012; Lempert 2014; Monani 2016). Consider the work of Salma Monani in her analysis of Danis Goulet's (2013) science fiction short film *Wakening*.

The sci-fi/horror movie is set in a dystopian time in which a colonizing group, the occupiers, have destroyed the environment and make it illegal for anyone else to possess land. Several protagonists emerge in this dystopia, the first being Weesageechak, a longstanding Cree trickster portrayed as a contemporary warrior woman in the film armed with archery equipment and protective medicine. She enters a theater in which people who once were captivated by the images on the stage or screen are now gone, with the few remaining asking to be saved from death. The initial reason for this dystopia is the violent actions of the other protagonist, Weetigo, a legendary Cree monster, who is portrayed as a forest elk hybrid creature who lives in the theater and is initially seen as the cause of the suffering. Yet Weesageechak, in seeking Weetigo in the theater, says that the occupiers have tricked Weetigo into being so destructive, and that it is the occupiers who are more powerful, Weetigo now being forgotten. Weetigo eventually turns away from ensnaring and killing Weesageechak and kills two occupiers who are about to kill a person. The film ends with both protagonists staring into each other with the noise of the occupiers in the background, as Weetigo disappears and Weesageechak stares into a brighter horizon with a wistful look.

In her interviews with Goulet, Monani (2016) discusses how the struggle of the protagonists arises from Cree storytelling. Goulet sets this story in the dystopian times of the occupiers. In the film, the protagonists are women and non-humans who have to figure out how to relate to each other again to resist the genocide and environmental destruction of the occupiers who are the true force of destruction and injustice. Both protagonists occupy social identities that are disrespected or villainized in Canadian or US settler colonialism, whether owing to gender, Indigeneity, or being nonhuman. The film emphasizes and honors the positive agencies of Weesageechak and Weetigo. In this sense, Weetigo is not entirely anthropomorphized and acts according to an agency that humans cannot fully comprehend or control but must respect. The film expresses Weesageechak's responsibility to respect and confront Weetigo and Weetigo's responsibility not to be fooled by the occupiers. Of course, the solution to surviving the dystopia lies in the reciprocal responsibility of both protagonists to work together in ways that honor each other. One way of interpreting *Wakening* is as an unfolding narrative of dialogue with ancestors and descendants, where what becomes apparent is the importance of reestablishing a relationship of reciprocal responsibility between the two protagonists, and emphasizing gendered and nonhuman agencies (see also Nelson 2013 for another example of this type of narrative relating to climate change).

In her analysis of Indigenous science fiction, gender, and futurism, Danika Medak-Saltzman (2017) writes, “Indigenous futurist work can and does also explore a variety of dystopian possibilities, which allows for critical contemplation about the dangerous ‘what ifs’ we might face and, more pragmatically, can aid us in our efforts to imagine our way out of our present dystopic moment to call forth better futures” (143). Medak-Saltzman focuses on how Indigenous science fiction works empower women and nonhuman protagonists. Looking at Nanobah Becker’s (2012) *The 6th World*, a futuristic film about the Navajo Nation working with the Omnicorn Corporation to create a colony on Mars, Saltzman-Medak claims that “it is women who are endowed with the ability to usher forth our collective futures, but it does so in a manner that complicates this notion and delinks it from being understood only through the lens of biological reproduction . . . [expanding] women’s roles and value beyond the limits imposed by patriarchy, colonization, and heteronormativity” (163). The film also brings out the protagonist agency of Navajo traditional corn, which plays multiple roles in the film through its spirituality, place in Navajo cultural heritage, association with sound scientific knowledge, and motivational value for imagining better futures (Medak-Saltzman 2017). Thus, *The 6th World* follows a long tradition of Indigenous science fiction that “promotes deeper understandings of biodiversity, cultural diversity, and refugia” (Adamson 2016, 219).

The short stories contained in *Love Beyond Body, Space, and Time: An LGBT and Two-Spirit Sci-Fi Anthology* (Nicholson 2016) further both these ends and more. Contributor Grace Dillon (2016) understands this anthology to be about “persistence, adaptation, and flourishing in the future, in sometimes subtle but always important contrast to mere survival” (9). These are what Gerald Vizenor (2008) calls native survivance stories. He explains:

> The native stories of survivance create active presence, more than the instincts of survival, function, or subsistence. Native stories are the sources of survivance, the comprehension and empathies of natural reason, tragic wisdom, and the provenance of new literary studies. Native stories of survivance are prompted by natural reason, by a consciousness and sense of incontestable presence that arises from experience in the natural world, by the turn of seasons, by sudden storms, by migrations of cranes, by the ventures of tender lady’s slippers, by change of moths overnight, by unruly mosquitoes, and by the favor of spirits in the water, rimy sumac, wild rice, thunder in the ice, bear, beaver, and faces in the stone. (11)

That survivance is curated here through two-spirit love stories makes their science fiction futurities that much more powerful.

8. Conclusion

Science fiction narratives such as those explored here are innovatively philosophical in their engagement with technologies as systems. Their imaginative mirroring of climate change—and other drivers of the so-called Anthropocene—in particular represent (re) descriptive analyses of technological systems. Moreover, the articulation of these technological systems is laden with careful visions of the future. Whereas the futurism of the philosophy of technology has focused on perceptions of risk or concerns about perils on the horizon, the literatures we have described offer diverse philosophical formulations of futures and the roles of/for technology therein. The futurities generated by these narratives accept the weight of past/present endings without being defined by them. Cultivating nonlinear and pluralistic temporalities, visionary science fiction frames technological systems holistically and contextually. In these worlds, the relationship between technologies and climate change (analogs) refuses reductive descriptions such as genesis and savior.

Visionary narratives such as these also have much to offer the philosophy of technology insofar as they are helpful for moving the literature beyond the Anthropocene discourse and colonial logics. "In a perilously warming world," Kate Rigby (2015, 2) writes, "the kinds of stories that we tell about ourselves and our relations with one another, as well as with nonhuman others and our volatile environment, will shape how we prepare for, respond to, and recover from increasingly frequent and, for the communities affected, frequently unfamiliar forms of eco-catastrophe." Toward these ends, (post) apocalyptic Indigenous, Afrofuturist, and/or feminist science fiction narratives are invaluable for their ability to frame environmental injustice intersectionally and (re)imagine just worlds. They do so by carefully attending to the intersections of gender, race, class, sexuality, etc. and the politics and technologies that produce and are produced by them (Otto 2012; Gaard 2014; Anson 2017). In addition to centering positionalities and identities too rarely encountered in mainstream science fiction, such stories work skillfully to thoroughly contextualize these "atypical" characters and their narrative perspectives. Even when the (post)apocalyptic conditions these characters experience do not mirror climate change explicitly or even metaphorically, their worlds and stories can be helpful so long as the anthropogenic causal mechanism and injustice of these breaking points remain central (Schatz 2012; Rigby 2015; Anson 2017). Regardless of the precise mechanism(s), the results are the same. If the characters and communities in these worlds cannot go backward, they must go forward.

Rather than looming on the horizon, here apocalypse occupies the present and, especially for post-apocalyptic worlds, the past. As "a moment of grave danger that also harbors liberating potentials," apocalypse is not The End but an ending, which, although tragic, offers the possibility of positive radical transformation (Dillon 2012; Rigby 2015). That the end of the world is already well under way only enhances these stories' moral/political applicability, for they imaginatively mirror how, for many peoples, environmental dystopia–apocalypse is hardly a new phenomenon (Whyte 2017). It is not only the temporal orientation of these worlds, but also whose futures are envisioned that set them apart. Indigenous, Afrofuturist, and feminist visionary narratives intentionally (re)center those on the receiving end of climate change and intersecting injustices. Rather than envisioning how those most responsible might redeem themselves or survive, these narratives refuse to reassure the privileged that their futures are secure. Quite the opposite, the stories—at their most radical and hopeful—reveal how privileged futurities must "give way" in both the stories themselves and the world beyond the page or screen (Vizenor 2008; Tuck and Yang 2012). The primary narrative arc, however, does not typically revolve around competing or incommensurable futurities but around conflicts internal to (re)imagining oppressed and marginalized futurities (Vizenor 2008; Dillon 2012). Instead, here we have characters and communities navigating the temporally, ecologically, and politically fraught (post)apocalyptic landscape by moving forward on their own terms (Vizenor 2008).

Thus, visionary science fiction works to (re)describe the present and past, as well as to (re)imagine the future. By engaging with visionary fiction, the philosophy of technology can refocus its efforts from pulling "us" back from the brink to initiating transformative climate justice moving forward. Both these strategies are necessary for departing from Anthropocene discourse so as to better align philosophy of technology with anticolonial, anti-racist, and feminist approaches to justice. Although surely there are many reasons that the stories we tell about climate change matter, their ability to resist, undermine, and propose alternatives to master narratives of technology associated with the Anthropocene must be counted among them.

Notes

1. Together Brown and Imarisha edited a volume of science fiction short stories written by social justice organizers. Imarisha writes, "All organizing is science fiction. Organizers and activists dedicate their lives to creating and envisioning another world, or many other worlds" (Brown and Imarisha 2015, 3).
2. To clarify, Selver is speaking of himself here as a 'god,' not Davidson, Lyubov, or humanity collectively.
3. The majority of text and analysis in this section is adapted from Whyte (2018), including many identical sentences. Whyte is an author of this chapter too. It would have been unnecessary to take pains to avoid repetition between the 2018 article and this chapter given that the section does not constitute the major contribution of this chapter.

References

Adamson, J. 2016. "Collected Things with Names like Mother Corn: Native North American Speculative Fiction and Film." In *Routledge Companion to the Environmental Humanities*, edited by U. K. Heise, J. Christensen, and M. Neiman, 2016–2225. New York, NY: Routledge.

Anson, April. 2017. "American Apocalypse: The Whitewashing Genre of Settler Colonialism." *Academia.edu*: 1–10. https://www.academia.edu/34949190/American_Apocalypse_The_Whitewashing_Genre_of_Settler_Colonialism (accessed August 20, 2020).

Aronofsky, Darren, dir. 2014. *Noah*. Paramount Pictures.

Asveld, Lotte, and Sabine Roeser, eds. 2012. *The Ethics of Technological Risk*. New York, NY: Routledge.

Atwood, Margaret. 2003–2013. *MaddAddam* trilogy. New York, NY: Random House, Inc.

Bay, Michael, dir. 1998. *Armageddon*. Touchstone Pictures.

Becker, Nanobah, dir. 2012. *The 6th World*. Independent Television Service.

Borgmann, Albert. 1987. *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago, IL: University of Chicago Press.

Brown, Adrienne Maree, and Walidah Imarisha, eds. 2015. *Octavia's Brood: Science Fiction Stories from Social Justice Movements*. Oakland, CA: AK Press.

Butler, Octavia. 1993/1998. *Earthseed* series. New York, NY: Warner Books, Inc.

Cafaro Phillip, and Richard Primack. 2014. "Species Extinction is a Great Moral Wrong." *Biological Conservation* 170: 1–2.

Cameron, James, dir. 2009. *Avatar*. 20th Century Fox.

Cuomo, Chris. 1998. *Feminism and Ecological Communities: An Ethic of Flourishing*. London, UK: Routledge.

Cuomo, Chris. 2011. "Climate Change, Vulnerability, and Responsibility." *Hypatia* 26: 690–714.

Cuomo, Chris. 2014. "Who Is the 'Anthro' in the Anthropocene." University of Georgia, Athens, GA.

Davis, Heather, and Zoe Todd. 2017. "On the Importance of a Date, or, Decolonizing the Anthropocene." *ACME: An International Journal for Critical Geographies* 16: 761–780.

Dillon, Grace, ed. 2012. *Walking the Clouds: An Anthology of Indigenous Science Fiction*. Tucson, AZ: University of Arizona Press.

Dillon, Grace. 2016. "Beyond the Grim Dust of What Was." In *Love Beyond Body, Space, and Time: An LGBT and Two-Spirit Sci-Fi Anthology*, edited by Hope Nicholson, 9–11. Winnipeg, MB: Bedside Press.

Emmerich, Roland, dir. 2004. *The Day after Tomorrow*. 20th Century Fox.

Feenberg, Andrew. 1999. *Questioning Technology*. New York, NY: Routledge.

Floridi, Luciano. 2014. "Technoscience and Ethics Foresight." *Philosophy & Technology* 27, no. 4: 499–501.

Gaard, Greta. 2014. "What's the Story? Competing Narratives of Climate Change and Climate Justice." *Forum for World Literature Studies* 6: 272–291.

Gabriel, Mike, and Eric Goldberg, dirs. 1995. *Pocahontas*. Walt Disney Pictures.

Garrard, Greg. 2012. "Part 5: Apocalypse." In *Ecocriticism: The New Critical Idiom*, 93–116. New York, NY: Routledge.

Goulet, Danis, dir. 2013. *Wakening*. ViDDYWELL Films.

Gross, Larry. 2014. *Anishinaabe Ways of Knowing and Being*. New York, NY: Routledge.

Grosvenor, Charles, dir. 1993. *Once Upon a Forest*. 20th Century Fox.

Haraway, Donna. 2008. *When Species Meet*. Minneapolis, MN: University of Minnesota Press.

Haraway, Donna. 2015. "Anthropocene, Capitalocene, Plantationocene, Chthulucene: Making Kin." *Environmental Humanities* 6: 159–165.

Herbert, Frank. *Dune*. 1965. New York, NY: Penguin Random House LLC.

Hopkinson, Nalo. 2000. *Midnight Robber*. New York, NY: Warner Books, Inc.

Hurley, Kameron. 2017. *The Stars are Legion*. New York, NY: Sage Press.

Ihde, Don. 1990. *Technology and the Lifeworld: From Garden to Earth*. Bloomington, IN: University of Indiana Press.

Jemisin, N. K. 2015/2016/2017. *The Broken Earth Trilogy*. London, UK: Orbit Books.

Johansen, Erika. 2014/2015/2016. *The Queen of the Tearling* series. New York, NY: Harper Collins Publishers.

Kareiva, Peter, and Michelle Marvier. 2012. "What Is Conservation Science?" *BioScience* 62: 962–969.

Kolbert, Elizabeth. 2010. "The Anthropocene Debate: Marking Humanity's Impact." *Yale Environment* 360. May 17, 2010. Web Dec. 17, 2015.

Kroyer, Bill, dir. 1992. *FernGully: The Last Rainforest*. 20th Century Fox.

Kubrick, Stanley, dir. 1964. *Doctor Strangelove or: How I Learned to Stop Worrying and Love the Bomb*. Columbia Pictures.

Le Guin, Ursula. 1976a. *The Word for World Is Forest*. New York, NY: Berkley Books.

Le Guin, Ursula. 1976b. "Introduction." In *The Left Hand of Darkness*. New York, NY: Penguin Random House LLC.

Le Guin, Ursula. 2004. *The Wave in the Mind: Talks and Essays on the Writer, the Reader, and the Imagination*. Boston, MA: Shambhala Publications, Inc.

Lempert, W. 2014. "Decolonizing Encounters of the Third Kind: Alternative Futuring in Native Science Fiction Film." *Visual Anthropology Review* 30: 164–176.

Lepori, Matthew. 2015. "There Is No Anthropocene: Climate Change, Species-Talk, and Political Economy." *Telos* 172: 103–124.

Lewis, Simon, and Mark Maslin. 2015. "Defining the Anthropocene." *Nature* 519: 171–180. Web. Dec. 17, 2016.

Little, Judith, ed. 2007. *Feminist Philosophy and Science Fiction: Utopias and Dystopias*. Amherst, NY: Prometheus Books.

Lloyd, Saci. 2009. *The Carbon Diaries*. New York, NY: Holiday House.

Medak-Saltzman, D. 2017. "Coming to You from the Indigenous Future: Native Women, Speculative Film Shorts, and the Art of the Possible." *Studies in American Indian Literatures* 29: 139–171.

Miller, George, dir. 2015. *Mad Max: Fury Road*. Warner Bros.

Mitchell, Audra. 2015. "Decolonising the Anthropocene." https://worldlyir.wordpress.com/2015/03/17/decolonising-the-anthropocene/

Mitchell, David. 2014. *The Bone Clocks*. New York, NY: Random House.

Monani, S. 2016. "Feeling and Healing Eco-Social Catastrophe: The 'Horrific' Slipstream of Danis Goulet's Wakening." *Paradoxa* 28: 192–213.

Morgan, Kass. 2013–2016. *The 100* Series. New York, NY: Little, Brown Books.

Morton, Timothy. 2013. *Hyperobjects: Philosophy and Ecology after the End of the World*. Minneapolis, MN: University of Minnesota Press.

Nelson, M. 2013. "The Hydromythology of the Anishinaabeg." In *Centering Anishinaabeg Studies: Understanding the World through Stories*, edited by J. Doerfler, N. J. Sinclair, and H. K. Stark, 213–233. East Lansing, MI: MSU Press.

Nicholson, Hope, ed. 2016. *Love Beyond Body, Space, and Time: An LGBT and Two-Spirit Sci-Fi Anthology*. Winnipeg, MB: Bedside Press.

Nolan, Christopher, dir. 2014. *Interstellar*. Paramount Pictures.

O'Brien, Robert C. *Z for Zachariah*. 1974. New York, NY: Simon and Schuster.

Otto, Eric. 2012. *Green Speculations: Science Fiction and Transformative Environmentalism*. Columbus, OH: Ohio State University Press.

Plumwood, Val. 1993. *Feminism and the Mastery of Nature*. New York, NY: Routledge.

Rigby, Kate. 2015. *Dancing with Disaster: Environmental Histories, Narratives, and Ethics for Perilous Times*. Charlottesville, VA: University of Virginia Press.

Risling Baldy, Cutcha. 2014, April 24. "Why I Teach 'The Walking Dead' in My Native Studies Classes." *The Nerds of Color*. https://thenerdsofcolor.org/2014/04/24/why-i-teach-the-walking-dead-in-my-native-studies-classes/

Robinson, Kim Stanley. 2004, 2005, 2007. *The Science in the Capital* series. New York, NY: Bantam Dell.

Rockstrom, Johan et al. 2009. "A Safe Operating Space for Humanity." *Nature* 461: 472–475.

Roddenberry, Gene. 1966–1969. *Star Trek, the Original Series*. Norway Productions, Desilu Productions, and Paramount Television.

Roddenberry, Gene. 1987–1994. *Star Trek, the Next Generation*. Paramount Pictures.

Sandler, Ronald. 2014. "The Ethics of Reviving Long Extinct Species." *Conservation Biology* 28: 354–360.

Schaffner, Franklin, dir. 1968. *Planet of the Apes*. 20th Century Fox.

Schatz, J. L. 2012. "The Importance of Apocalypse: The Value of End-of-the-World Politics while Advancing Ecocriticism." *Journal of Ecocriticism* 4: 20–33.

Scott, C. 2016. "(Indigenous) Place and Time as Formal Strategy: Healing Immanent Crisis in the Dystopias of Eden Robinson and Richard Van Camp." *Extrapolation* 57: 73–93.

Dr. Seuss. 1972. *The Lorax*. New York, NY: Random House.

Shrader-Frechette, Kristin. 2002. Environmental Justice: Creating Equality, Reclaiming Democracy. Oxford, UK: Oxford University of Press.

Sprague, Lee. 2017. Personal Discussion, August 6.

Traub, Courtney. 2018. "From the Grotesque to Nuclear-Age Precedents: The Modes and Meanings of Cli-fi Humor." *Studies in the Novel* 50: 86–107.

Thompson, Allen, and Jeremy Bendik-Keymer. 2012. *Ethical Adaptation to Climate Change: Human Virtues of the Future*. Cambridge, MA: MIT Press.

Thompson, Paul. 2010. *The Agrarian Vision: Sustainability and Environmental Ethics*. Lexington, KY: University Press of Kentucky.

Tuck, Eve, and K. Wayne Yang. 2012. "Decolonization Is Not a Metaphor." *Decolonization: Indigeneity, Education, & Society* 1: 1–40.

Ullrich, J. K. 2015. "Climate Fiction: Can Books Save the Planet." *The Atlantic*. Aug. 14.

Vaidyanathan, Gayathri. 2014. "Can Humans and Nature Coexist? Conservationists Go to War over Whether Humans Are the Measure of Nature's Value." *Scientific American* 10. Web Sep. 1, 2015.

Vizenor, Gerald, ed. 2008. *Survivance: Narratives of Native Presence*. Lincoln, NE: University of Nebraska Press.

Vogel, Steven. 2015. *Thinking Like a Mall: Environmental Philosophy after the End of Nature*. Cambridge, MA: MIT University Press.

Whyte, Kyle Powys. 2017. "Our Ancestors' Dystopia Now. Indigenous Conservation and the Anthropocene. Routledge." In *Companion to the Environmental Humanities*, eds. U. Heise, J. Christensen, and M. Niemann, 206–218. Routledge.

Whyte, Kyle Powys. 2018. "Indigenous Science (Fiction) for the Anthropocene: Ancestral Dystopias and Fantasies of Climate Crises." *Environment & Planning E: Nature and Space* 1 (1–2): 224–242.

Yusoff, Kathryn 2018. *A Billion Black Anthropocenes or None*. Minneapolis, MN: University of Minnesota Press.

CHAPTER 25

A FRAMEWORK FOR THAWING VALUE CONFLICTS IN THE GMO DEBATE

SAMANTHA NOLL

1. Introduction

This chapter explores the ethical dimensions of one of the most contentious applications of agricultural biotechnology today: the genetic modification of food crops and animal breeds. The debate surrounding the application of genomics technology to food is dangerously polarized (Rich n.d.; Tester 2001; Thompson 1993) and continues to be actively discussed in the public sphere (Maghari and Ardekani 2011). Supporters of genetic modification often argue that it has many advantages, as genetically modified organisms (GMOs) can be designed to reduce reliance on pesticides and herbicides, improve disease resistance and nutritional content, and increase crop yields and quality etc. (Thompson 2006; Toft 2012). In reply, critics often argue that GMOs should not be used as they negatively impact the environment and/or animal welfare, compromise consumer health, promote the exploitation of farmers, and increase overall food risk. These competing views, and the subsequent sense of crisis, has intensified as new GMOs enter the market.

The first section of this chapter uses the AquAdvantage salmon debate to highlight the most common arguments made concerning the adoption of GMOs. It then goes on to break down these arguments into the following five categories of concern: impacts to (1) individuals, (2) society, (3) the environment, (4) animal welfare, and (5) general ontological concerns. This analysis teases out the value positions that provide justification for these concerns, illustrating how the polarization of the public GM debate stems from normative conflicts, as Thompson (2006) argues, rather than a lack of empirical research. It then identifies two barriers to achieving consensus concerning GMOs, the problem of normative freezing and the problem of ontological inflexibility.

After weighing and dismissing mandatory labeling as one possible solution, the chapter introduces the "GMO Value Framework" as a reflexive approach to help cultivate fruitful value-focused discussions with the aim of mitigating conflict. Specifically, the framework expands Beauchamp and Childress' (2001) principlist ethic to include additional environmental and animal welfare focused sub-principles, creating a matrix that aligns with the five categories of concerns. The chapter ends by utilizing this framework to analyze the AquAdvantage salmon debate, illustrating its usefulness for providing a shared terminology and for identifying substantive concerns regarding individual genomic applications, which is necessary for conflict management.

This chapter adds to the literature in philosophy of technology and bioethics on managing value and policy conflicts, which increasingly arise from advances in modern biotechnology. As this chapter provides an overview of the most prominent concerns raised by the public when discussing GMOs, the chapter is intended to assist researchers and public officials working in this area of biotechnology. It also provides analytical tools that may help philosophers, policymakers, and scientists to begin unfreezing the GMO debate and to facilitate future discussions concerning how biotechnologies should be applied. It should be noted here that this analysis largely focuses on the GM debate in the context of the United States and Canada. Community discussions concerning GM foodstuffs is a global phenomenon, with other countries taking a wide range of regulatory positions concerning these products. For example, in 2002, China required that five distinct types of GM crops (soybean, corn, rapeseed, and tomatoes) be labeled (Zhao et al. 2019). Similarly, Europe has some of the strictest regulations concerning GM products in the world (Davison and Bertheau 2010). In this context, the public demanded that they be given information necessary for making informed food-choices. As a result, the European Union adopted legislation ensuring that GM products are traceable, including mandatory labeling of food products and animal feed that contain GMOs. It is the author's hope that the analysis in this chapter will help to enrich future work on managing value and policy conflicts in these and other contexts beyond the United States and Canada. However, before presenting this analysis, it is important to establish how genetic modification is defined in this chapter.

Genetic modification (GM), or genetic engineering, roughly signifies the modification of genes to allow for a change to be passed onto future generations (Bawa and Anilakumar 2013) or the transference of selected individual genes from one organism to another, be those of the same species or between species (Phillips 2008). This definition is broad, as it includes everything from selective breeding, where genes are modified by breeding two animals/plants of the same species with desired traits, to the relatively new activity of inserting genetic material from one species into another species (Savulescu 2011). Thus, genetic modification acts as an umbrella term, as it includes traditional breeding programs, as well as new biotechnologies, such as the development of clustered regularly interspaced short palindromic repeats or CRISPR (Han and She 2017). Now that we have clarified the definition of genetic modification, the next section of this chapter presents a case study and analysis to ground the GMO framework.

2. AquAdvantage Salmon: A Case-Study

As a point of reference for the ethical analysis of genetically modified organisms, the AquAdvantage salmon debate can help to ground our theoretical analysis (Gallegos 2017). This debate is particularly contentious and helps to illuminate various ethical issues associated with the adoption of genetically modified organisms, as the prospect of GM salmon becoming available in the United States and Canada sparked a prolonged public discussion of the potential benefits and harms of this product (Waltz 2017). While there are many genetically modified plant-based projects in the food supply, GM salmon is distinct, as it is one of the first GM animals to be approved for sale in Canada and by the United States Food and Drug Administration (FDA) (Grossman 2016). After two decades of public debate, this product was approved for sale in Canada in 2016, with over 4.5 tons sold in 2018 alone (Kassam 2016). As such, the subsequent discussion includes concerns aimed at genetic modification in general, and genetically modified animals in particular.

Specifically, the public debate concerns whether AquAdvantage salmon should be sold for consumption. This salmon is “a fast-growing transgenic fish containing a gene encoding Chinook salmon growth hormone under the control of an antifreeze protein promoter and terminator from ocean pout” (Eenennaam and Muir 2011). The main benefit of these changes is that the fish grows to market size in 16 to 18 months, in contrast to the three years that it takes non-modified salmon. It is one of the first genetically engineered animals to be used for food consumption and AquAdvantage salmon has undergone one of the most exhaustive regulatory assessments in history by the FDA and Health Canada and the Canadian Food Inspection Agency. Concerning FDA approval, Eenennaam, and Muir (2011) state that “if the GE animal is intended as a source of food, as is the case with the AquAdvantage salmon, FDA assesses whether the composition of edible tissues differs and whether its products pose more of an allergenicity risk than non-GE counterparts” (706). The FDA also requires environmental assessments of the animal and of the facilities where they will be raised. Finally, the FDA determines whether or not the company’s claims hold. In this case, it will be determined if the GM salmon grows faster than traditional or selectively bred salmon. The producers of AquAdvantage salmon initially applied for approval in the early 1990s and, after a thorough assessment, it was finally approved for sale in November 2015 (Grossman 2016). Canada soon followed suit, approving this GM product for sale in 2016, after four years of safety testing (Kassam 2016).

While the producers of the salmon argue that it is a safe and more sustainable alternative to non-modified counterparts, both environmental and consumer safety groups, including Earthjustice and the Center for Food Safety (Doezema 2017), have adamantly voiced concerns, often calling for the removal of the product from the US market (Gallegos 2017). In fact, genetically modified food products are currently receiving strong pushback due to a wide range of worries, such as a lack of trust in the new technology, possible negative health and environmental risks, and the novel nature of transgenic animals (Thompson 1997; Rollin 2015). As Marris (2001) argues, "the anti-GMO lobby accuses proponents of this technology of pushing the introduction of GMOs into agriculture without adequately considering health and environmental risks" (545). Concerning GM salmon, consumer groups are also concerned with transgenic applications, or moving genetic material into unrelated organisms (Mather et al. 2012; Savulescu 2011). This led anti-GMO advocates to give AquAdvantage salmon the label "frankenfish," implying that it is ontologically "unnatural" (Goldschmidt 2015). In response, supporters of GM technology argue that tribalism, distrust of corporations, and a lack of scientific literacy (Ronald 2016) caused the anti-GMO lobby to blow the potential risks of salmon out of proportion (Marris 2001). AquAdvantage salmon was ultimately approved for sale in the United States in 2015 (Grossman 2016). However, environmental and consumer groups continue to call for its removal from the market, with Earthjustice and the Center for Food Safety suing the FDA over their approval of the product (Doezema 2017).

3. An Analysis of the GM Debate

The public debate over AquAdvantage salmon highlights key arguments concerning the products of GM technologies. These can be placed into the following five categories of concern: Impacts to (1) individuals, (2) society, (3) the environment, (4) animal welfare, and (5) ontological concerns (Stirling and Mayer 2001).[1] Impacts to individuals include concerns over human health risks, nutritional value, consumer choice etc. In contrast, the environmental and social categories encapsulate wider impacts at the systems level, including impacts to ecological systems and the distribution of harms and benefits among societal groups. The animal welfare category includes impacts to animals, as these play an important role in public discussion concerning GMOs (Thompson 1997; Rollin 2015). In the agricultural context, biotechnology could be used to increase animal growth rates and/or tolerance to cold and dehydration, with possible increases of animal suffering (Fox 1992; Thompson 1997). Finally, the ontological category captures potential impacts to species type and arguments concerning transgenic modifications or those that violate species boundaries (Macnaghten 2004; Noll 2013). These are important to include, as such worries often motivate conceptual or category-based critiques regarding genetic modifications that move beyond individual, social, and environmental impacts. It should be noted that the five categories are meant to provide a brief overview of positions, rather than an exhaustive list. They are, by design, rough and thus should not be taken as absolutes. Rather, their purpose is to highlight key positions taken in the debate and to provide conceptual clarity.

Arguments in favor of approving genetically modified salmon for consumption tend to fall into four of the five categories outlined above. First, supporters claim that adopting AquAdvantage salmon could decrease the cost of purchasing salmon for individual consumers (United States Food and Drug Administration 2019). Second, they marshal arguments regarding wider social impacts, such as providing jobs for local communities and increasing food availability (or food security). Third, supporters discuss environmental benefits, such as how GM salmon could provide a sustainable alternative to fisheries that rely on native fish populations and potentially reduce negative environmental impacts that result from over-harvesting. Fourth, supporters also discuss general ontological worries when they state that GM salmon are virtually identical to non-GM species in all ways that matter for food production.

Likewise, arguments marshaled against adopting GM salmon include a wide range of concerns that highlight several of the most prevalent positions taken in the GM debate. For example, common individual-focused arguments against adopting genetically modified salmon include claims that they (1) are not safe for consumers, (2) push technology onto people without their consent (i.e., they are not labeled), and (3) are not cost effective (Thompson 2006; Sandler 2005). Other critics are worried about the wider environmental harms that could occur if GM salmon escape into the environment, as this could impact the health of existing fish populations and/or cause wider harms to ecosystems (Marris 2001). As such, these concerns encapsulate individual, environmental, and animal welfare impacts. Opponents are also troubled by the "unnaturalness" of the GM salmon, or impacts that stem from the ontological status of a being that contains genetic material from two or more unrelated organisms (Rollin 2015; Savulescu 2011). These types of arguments critique the scientific processes used in the development of the GMO and/or begin with implicit ontological assumptions concerning types of beings and the boundaries between lifeforms and whether we should "play god" (Rollin 2015; Savulescu 2011).

In addition to providing clear examples of the types of arguments deployed for and against GMO technology, the analysis of AquAdvantage salmon highlights an important aspect of the GMO debate: specifically, that arguments made in support of, or against, adopting genetically modified products are predominantly guided by value commitments (Thompson 2006). According to Thompson (2006), normative stances guide most arguments concerning GM crops, even the marshaling of scientific evidence used to support one side or another. For instance, let us return to individual impacts. Arguments for or against GM products of this type can be reduced to the following basic structure: (1) GMOs have *x* properties. (2) These properties are harmful or beneficial for *y*. (3) Thus, we should use or discontinue the use of GMOs. Here the leap from 2 to 3 entails that *y* has value of some sort, be that intrinsic or instrumental. For instance, arguments referencing health impacts often make use of the hidden or unstated premise and conclusion that human beings are intrinsically valuable and thus should not be harmed. If consumers were not valuable, then negative impacts would be treated as empirical facts and not as reasons for changing behavior or discontinuing use.

Additionally, the leap from 2 to 3 often makes use of hidden premises regarding why we should care about these harms and benefits. By way of illustration, radiation is harmful for cancer cells but we do not then make the claim that radiation treatments are "bad" and the technology should be abandoned. In contrast, we prioritize the benefit to cancer patients over the harm to cancer cells in this instance. This is because we have robust normative frameworks encoded in our policies and social structures that provide basic (though contestable) justification for recognizing the ethical worth of a being and/or addressing conflicting normative claims. For instance, arguments concerning potential health impacts include empirical components, such as "facts" or probabilities concerning whether a product *is* harmful to humans, and normative components that justify why these risks *should* be minimized (Shrader-Frachette 2002). In addition, worries that GMOs violate consumer autonomy make such a normative leap explicit, as the right to self-governance is historically grounded in Kantian moral theory (Christman 2018). In short, how we value the impacted party directly impacts the conclusion of the argument.

Political values play a similar role guiding arguments regarding GMOs. For example, egalitarian frameworks are often used to justify claims that maintaining social goods should be prioritized (Thompson 2006; Noll 2017) or that resources should be distributed throughout society in a specific way (Rawls 2001; Shrader-Frachette 2002). While this term is contested in the wider philosophical literature, at its most fundamental level, egalitarianism begins with the basic claim that social goods, resources, and harms need to be distributed in an equitable fashion across the population (Arneson 2013). As agriculture both utilizes and creates resources, it is not surprising that concerns over what constitutes just distribution (of risks and benefits) often guide arguments in this context. For example, one could argue that the poor are likely to bear an inordinate and unjust share of the risks associated with GM products, as it is harder for them to opt out of purchasing these products due to the higher price of organic non-GM crops (Alkon 2014) or their non-availability, in the case of global food aid (Barnhill et al. 2016). Thus, we should either make it possible to opt out (through the use of labeling, government subsidies, etc.) or discontinue the use of GMOs.

Egalitarian paradigms also guide arguments regarding how ecological impacts should be distributed, as environmental philosophers have expanded normative frameworks to include ecosystems, biotic communities, and non-human animals (Noll 2017). For example, "deep ecologists (Naess 1973) and ecofeminists (Warren 2000) emphasize the view that the natural world does not exist solely for the use of humans, but accept a biocentric view of the natural world where it has intrinsic value" (Noll 2017, 30). Similarly, animal ethicists, such as Singer (2015), Regan (1995), and Palmer (2010) place non-human animals firmly in the ethical sphere. As biotechnologies can modify animal bodies and behaviors, animal ethics are often utilized when discussing the impacts of such technologies (Rollin 2015; Noll 2013; Savulescu 2011). Placing individual animals and ecosystems within the ethical sphere sets the stage for contentious discussions concerning what constitutes the just distribution of benefits and harms associated with biotechnology. For instance, concerns over the ecological impacts of "super weeds" are often guided by a) the basic normative commitment that environments are valuable and/or b) an egalitarian concern that the environment is bearing an unjust share of the risks associated with GM technology. Thus, arguments regarding GM technology include empirical, political, and normative components. From this position, it becomes clear that "facts" alone cannot be used to address public worries. It follows then that an important part of the analysis of GM debates needs to make explicit the value claims and/or normative frameworks being used.

4. Normative Freezing as a Barrier to Consensus

If deeper value commitments guide the most prominent arguments concerning GM products, then how can we move forward the increasingly polarized GMO debate in a fruitful way? How should we apply current biotechnologies? I argue that there are two potential barriers that will need to be addressed before we can attempt to find public consensus on this issue: (1) The problem of *normative freezing*; and (2) the problem of *ontological inflexibility*. First, if value commitments ultimately guide the use of facts, rather than being changed by facts, then deeper normative positions regarding GMOs could be uncompromising or inflexible. In this vein, Thompson argues that "parties on either side can continually shift the burden of proof to the other side with new empirical data . . . ," thus ensuring that the debate will continue until both parties are exhausted (Thompson 2006, 76–77). Thompson uses this insight to highlight how normative positions play a key role in biotechnology discussions. However, this insight also illustrates a general reluctance to reweigh and reevaluate positions when new evidence is presented, or the problem of 'normative freezing.' This reluctance could act as a barrier to obtaining consensus on GMOs, as improving scientific literacy or providing more information to the public may not address underlying normative concerns.

In addition, normative commitments can conflict, and myopically focusing on providing more factual information may not address these conflicts. There is a large body of work in environmental philosophy that explores how different ethical approaches provide important ethical guidance, but are ultimately incompatible with one another (Callicott 1980; Jamieson 2008). When analyzing the GM debate, a similar tension is found between those approaches focusing on impacts to humans, non-human animals, and social and ecosystem impacts. Focusing on one of these areas could provide guidance that ultimately conflicts with actions that could benefit another area. For example, providing GM food aid may harm farmer livelihoods in a specific context, but improve overall food security in that community. Relative valuations of impacts to human individuals, social groups, non-humans, and the environment cut across arguments marshaled by both supporters and critics of GM technology. Applying a single ethical approach then, would not bring consensus even within the already calcified camps, let alone build bridges between these two groups. Thus, ethical positions, grounded in different valuations, could end in freezing or hardening the dispute to the point where consensus is impossible. This is what is meant by the problem of normative freezing.

Second, there is the problem of ontological inflexibility. Ontological concerns often arise when discussing GMOs (Savulescu 2011), as genetic changes could call into the question the nature and/or properties of biological entities. Current biotechnologies open up a plethora of possibilities that could violate species boundaries, as we move genetic material from one species (such as an African toad) to another species (such as a papaya). AquAdvantage salmon is the product of transgenic applications of biotechnology, as it contains genetic material from unrelated organisms, specifically the Atlantic salmon and ocean pout (Mather et al. 2012). These new biotechnologies and cross-species applications have the effect of bringing assumptions concerning properties of species to the forefront of the GM debate. Such concerns are "ontological," as they grapple with determining the type or kinds of entities that exist; their qualities, properties, or attributes (i.e., "naturalness"); and the basic nature of what it means to be a member of a species (Inwagen and Sullivan 2019; Savulescu 2011).

While it is important to recognize these concerns, as they play a role in the debate, I argue that the problem of ontological inflexibility may often be reduced to the problem of normative freezing, as ontological claims can be guided by normative worries. For example, both supporters and critics of GMOs seem to be making an ontological argument, when they appeal to the "naturalness" of the plant or animal. However, this term could capture a wide range of concerns, beyond determining the biological properties or attributes of food products. According to Chambers et al. (2018), there is no legal definition of the term "natural" and for this reason consumers have formed their own opinions as to what constitutes this property or attribute. This position is corroborated by an analysis of consumer comments sent to the FDA. According to Dominick et al. (2017), "80% of respondents thought that 'all natural' products would mean no antibiotics, no hormones, and no preservatives added in processing, with 60% of respondents saying that natural food would have improved animal welfare practices, improved nutritional value, and improved food safety" (263). Various studies on "natural" food conceptions identified a plethora of definitions, ranging from healthiness, environmental impacts, how they are grown or processed, what inputs are used, and importantly for this chapter, if and how they are genetically modified (Dominick et al. 2017).

Complications surrounding the definition of "natural" illustrate how ontological claims regarding properties and categorizations could signify a number of arguments beyond those that are purely ontological. When critics of AquAdvantage salmon call them "frankenfish" (intimating that they are unnatural), they are making an ontological claim. However, they could also be making a wide range of implied normative claims concerning the safety, healthiness, environmental impact of this product. Thus, the problem of ontological inflexibility could, in some cases, actually act as a red herring argument, diverting attention away from the main barrier to coming to consensus, that of normative freezing. For this reason, ontological worries will not be included in the framework below.

5. GMO Labeling as an Incomplete Normative Strategy

Due to the complicated nature of weighing technological impacts, allowing individuals to opt-out of eating genetically modified food is one strategy that has been proposed to address the conflict. If you don't like GMOs for whatever reason, you can simply not buy them and thus not support this biotechnology with your dollars. This position falls in line with Thompson's (2006) argument that GMO products should be labeled so that consumers can make choices that align with their values. Mandatory labeling would honor an individual's ability to choose and to support farming practices that they care about. At face value, this may be a viable strategy to reduce tensions between those who desire (or are ambivalent about) genetically engineered products and those who do not. However, there are at least two critiques of this position that need to be explored.

First, GMO labeling may not be an adequate way to address normative concerns, as it simply pushes the debate back to the policy level. The labeling debate is currently one of the leading policy issues regarding biotechnology in the United States and Canada at both the national and state levels (Marchant and Cardineau 2013). In contrast to the European Union which has some of the strictest GM regulations and mandatory labeling requirements (Davison and Bertheau 2010), Canada does not require GMOs to be labeled (Government of Canada 2020). However, they must undergo the same safety assessments that are required of any food product introduced into the market. This position is highly controversial, with consumer groups actively lobbying for the adoption of mandatory GM labeling regulations. In the United States, the Food and Drug Administration (FDA) has refused to require that GM foods be labeled. However, the debate is not over, as it has largely moved to the state level, where several states are considering GM labeling legislation (Byrne et al. 2014; Marchant and Cardineau 2013). In addition to weighing the legality of labeling requirements, this debate is also guided by several of the normative concerns identified above, including respecting consumer choice, potential health impacts, and preventing environmental harm.

Second, labeling would allow consumers to "vote with their dollars," but may not effectively address wider social and environmental concerns, such as potential damage to ecosystems and social injustices associated with the distribution of risks and benefits. If a portion of consumers purchase food that could cause ecological damage, then this damage would continue (albeit on a potentially smaller scale), even if a sub-set of consumers decided not to support the practice. For this reason, while mandatory labeling may help alleviate some tension, it will not fully remove the need to address the problem of normative freezing.

6. The GMO Value Framework

If our goal is to cultivate fruitful value-focused discussions with the aim of mitigating conflict, then we need a reflexive framework that captures the five types of concerns discussed above, as these are at the heart of the value conflicts. Fortunately, we do not need to create such a framework from scratch, as we can combine ethical standards from biomedical and research ethics, environmental philosophy, and animal welfare science and apply them in this context. The foundation of the GMO Value Framework is a practical approach for decision making found in biomedical ethics called principlism (Beauchamp and Childress 2001). This approach condenses value concerns into principles or "rules of thumb" designed to give people from diverse backgrounds an easily grasped set of moral standards to help guide decision making (Beauchamp 1995; Beauchamp and Childress 2001). The basic principles highlight desirable values that are often accepted as important within societies (DeMarco 2005). Thus, it should be noted here that principlism offers a "common morality" ethical theory, as the premises are taken directly from "the morality shared in common by the members of a society—that is, unphilosophical common sense and tradition" (Jecker et al. 2007, 147). As such, Beauchamp and Childress (2001) are clear that this approach should not be considered a comprehensive ethical theory.

As genetic engineering applications proliferate, an expanded version of this flexible framework could help guide discussion of GMOs and related technologies, as stakeholders weigh ethical impacts. It is especially well-suited for this purpose as biomedical contexts are also fraught with value-based conflicts, where doctors, patients, family members, and other stakeholders must make difficult decisions. The biomedical framework typically includes four basic principles: (1) respect for autonomy (which requires that we respect individuals' decision-making capabilities), (2) beneficence (which requires that we prevent harm and provide benefits), (3) nonmaleficence (which requires that we do not cause harm to others), and (4) justice (which requires that benefits and risks be fairly distributed; Beauchamp 1995; Beauchamp and Childress 2001).

Using these principles as a starting point for conversation could help us to identify and clarify value commitments that fall within these parameters, as well as any value-based friction. For example, autonomy would mandate that we explore and take seriously concerns over whether or not consumers can opt out of eating genetically engineered foodstuff. Beneficence would require that we list potential benefits and ways that this application could limit harms to individuals, be those humans or animals. Nonmaleficence would ask us to determine how this modification could cause harm to others. Finally, justice would require that we explore how risks and benefits are distributed across a community and any negative effects to historically marginalized populations. Framing individual and social concerns in this way could also highlight how distribution issues may conflict with individual-focused concerns in a nuanced manner.

Additionally, particular attention should be paid to benefits and harms that impact animals and the environment, as both areas do not typically fall within the purview of principlism proper. As highlighted by the case-study above, these types of impacts can fall under the principles of beneficence and nonmaleficence, as the most prominent arguments concerning GM products revolve around potential benefits and harms to individual humans, non-human animals, and ecosystems. However, expanding the scope of principlism is a controversial view in the bioethics literature (Rollin 2012). For example, Walker (2006) argues that while the four principles can be applied to non-human animals, principlism is primarily utilized in human-centered clinical and research settings. The result is a disconnect between the structure of bioethical guidelines for humans and guidelines that pertain to non-human animals, especially in the United States. The principle of autonomy is understood as ethically crucial for humans, while welfare is prioritized in ethical guidelines for animals. Both Rollin (2012) and Walker (2006) argue that limiting the scope of principlism to human subjects helps researchers avoid questioning whether animal experimentation is consistent with the principles. Indeed, when they are applied uniformly, the majority of animal research (and animal agriculture) could be considered unethical (Rollin 2012). Additionally, the debate concerning whether the principle of autonomy should be applied to non-human animals is particularly contentious, as such applications tend to depend on research regarding species specific animal cognition. However, for the purpose of weighing potential benefits and harms associated with genomics applications, beneficence and nonmaleficence should be expanded to include impacts to the environment and non-human animals.

In particular, it is imperative that environmental effects be included as new biotechnologies could have negative influences on species integrity, biodiversity levels, and ecosystem functioning (Batie and Ervin 1999; Hails 2009; Thompson 2006). This is a particularly important category, as "data from the US Department of Agriculture (USDA) show that farmers intend to plant approximately 80 percent of soybean acreage, 70 percent of cotton, and 38 percent of corn to transgenic varieties" (USDA 2019). Due to their widespread use, these plants will likely move into ecosystems throughout the United States (Erwin and Welsh 2006). In this vein, agricultural biotechnology has been described "as a tsunami washing over agriculture—with fundamental impacts" to our food systems, economic markets, and environmental sustainability (Batie and Ervin 1999, 1). Batie and Ervin (1999) break down environmental concerns into three categories: (1) Pesticide use impacts, (2) Non-target species impacts, and (3) Pest and virus resistance. The first category focuses on changes in agricultural inputs and captures the reality that GMOs could change pesticide usage rates in environmentally beneficial or harmful ways. Second, there are concerns over whether certain species are being harmed by GMOs, as agricultural lands are habitat to a multiplicity of organisms, many of which are ecologically beneficial (Losey et al. 1999; Gianessi and Carpenter 1999). The loss of "beneficials," such as pollinators, butterflies, and insects eaten by birds, could lead to a loss of biodiversity in adjacent ecosystems, as well. The final category captures the worry that GMOs may become herbicide-resistant, and thus difficult to control, or transfer modifications to weed relatives (Hubbell and Welsh 1998; Linder and Schmidt 1995). While maintaining current levels of production partially underscores the third concern, the categories are more generally guided by worries that GMOs could harm ecosystem functioning, native species integrity, and biodiversity. Thus, we need to include at least these three sub-principles (under beneficence and nonmaleficence) when discussing the efficacy of GMOs. These will help stakeholders identify specific environmental concerns.

For this reason, sub-principles regarding animal welfare should also be included in our expanded framework. In contrast to the four principles, ensuring that animal welfare standards are respected and in place in research settings is required in most countries worldwide (Gilbert et al. 2012; Rollin 2012). Animal welfare laws, both in the United States and abroad, have helped to ensure that vertebrate animals used in research are given "humane" treatment guided by standards of care (Gilbert et al. 2012; Rollin 2006; Mellor 2017). Common ethical systems of analysis include the Five Freedoms (Webster 2016), the 3Rs (Fenwick et al. 2009), and the Five Domains (Mellor 2017). While each is unique, they all include mandates aimed at ensuring the bodily health (i.e., freedom from hunger and thirst), mental wellbeing (i.e., freedom from fear and distress), and respect for species typical behavior (i.e., freedom to carry out important patterns of behavior), as much as possible (Webster 2016). Drawing directly from these frameworks, animal welfare focused sub-principles should be included in our larger framework that encompass at least three specific areas: animal bodily integrity; quality of life; and respect for species typical behavior. These animal welfare sub-principles will help to highlight important impacts to any food animals that are genetically modified when discussing the ethical status of a genetic engineering application. Adding these sub-principles pushes us to assess animal welfare impacts, as we weigh human-centered individual and social concerns.

Thus, the GMO Value Framework should include the four core principles of (1) respect for autonomy, (2) beneficence, (3) nonmaleficence, and (4) justice (Beauchamp and Childress 2001); the environmental focused sub-principles of (5) ecosystem functioning, (6) native species integrity, and (7) maintaining biodiversity; and the animal welfare sub-principles of (8) animal bodily integrity; (9) quality of life; and (10) respect for species typical behavior (Table 25.1). Each of these principles and sub-principles is designed to highlight nuances found in the five categories of concern listed in the table that guide arguments in support and critical of adopting food products produced through genetic engineering (Stirling and Mayer 2001). Unlike in biomedical ethics, where each principle is used to guide behavior, the GMO Value Framework is designed to highlight different value commitments and concerns in order to help prompt more fruitful discussion and to help identify underlying conflicts. As new technological applications arise, this flexible framework can provide a shared terminology and value-focused starting point for conflict management. Each principle will need to be weighed and balanced with the others when identifying the optimal action

Table 25.1. Concerns, GMO Value Frameworks, and Value-Focused Questions

The Five Categories of Concern	The GMO Value Framework	Value-Focused Questions
Impacts to Individuals	1. Respect for Autonomy* 2. Beneficence 3. Nonmaleficence	Are decision-making capabilities being respected? What are the benefits of adopting this product? What are the potential harms?
Impacts to Society	4. Justice	How are the harms and benefits distributed across society?
Impacts to the Environment (Sub-Principles under Beneficence and Nonmaleficence)	5. Ecosystem Functioning 6. Native Species Integrity 7. Maintaining Biodiversity	What are the impacts to ecosystems? What are the impacts to native species? Are biodiversity levels being affected?
Impacts to Animal Welfare (Sub-Principles under Beneficence and Nonmaleficence)	8. Animal Bodily Integrity 9. Quality of Life 10. Respect for Species-Typical Behavior	What are the impacts to the body of the animal? Are any negative mental states produced? What are the changes to the behavior of the animal?

* The first four principles are taken from Beauchamp and Childress 2001.

concerning the new biotechnology. This process requires a conversation, be that at the national, state, or local level. It could also be used at the individual level, to help those unsure of new technologies identify their personal position. As every human consumes food products, such exploration is imperative for choosing a diet that supports their values.

7. The GMO Value Framework and Conflict Management

To illustrate the importance of identifying value positions and tensions, the framework needs to be connected to current work in conflict management. Agricultural extension specialists, who provide non-formal education to rural communities, have experience dealing with conflict between a wide range of stakeholders due to environmental scarcity, the degradation or depletion of resources, and concerns over unequal distribution of harms and benefits (Ahmadvand and Karami 2007). As agricultural practices, stakeholder interests and needs, and best resource management practices are constantly changing (Ahmadvand and Karami 2007; Owen et al. 2000), conflicts can rarely be resolved but need to be actively managed. Important to this work is the distinction between conflict "resolution," or the settlement of conflicts, where everyone's values and interests are satisfied, and "management," or containing value conflicts while working towards tangible improvements (Ahmadvand and Karami 2007; Walker and Daniels 1997). Complex conflicts can often never fully be resolved, as coming to agreement entails that value and interest incompatibilities are resolved, and this is difficult. Rather, complex conflicts guided by competing values, such as the debate over GMOs, need to be managed so that they do not become intractable or destructive.

An important strategy for conflict management is the Progress Triangle, which breaks conflicts down into three key dimensions (Walker 2019; Ahmadvand and Karami 2007). These interrelated dimensions are the procedural, relationship, and substantive aspects of the conflict. The relationship component includes examining the connections between the parties in the conflict, their history, and the trust and respect between the groups. In the context of the GMO debate, the lack of trust of companies that genetically modify food (Ronald 2016) is a prime example. The procedural dimension asks us to assess the decision-making structure, including policy, legislative, and regulatory bodies. These two dimensions are important; however, the substantive dimension is the main cause of conflicts: tangible issues that are marked by a normative component such as "righting a wrong" or helping to bring about a better future (Walker 2019). Questions guiding substantive exploration include: What are the issues? What is the source of tension between these issues? Are interpretations varied? Are there any synergies or mutual gain opportunities? While philosophy may not be able to provide as much insight concerning the first two dimensions, which seem best explored by the social sciences, philosophical exploration is key to understanding substantive issues. As such, when placed in this context, the GMO Value Framework above is designed to provide insights into the substantive barriers to conflict management and thus provide a necessary contribution to address normative freezing.

8. A Re-analysis of AquAdvantage Salmon

Using the principles in the GMO Value Framework for personal reflection and a starting point for conversation could be a productive way to help stakeholders identify the substantive issues at the heart of biotechnology conflicts, as well as their value commitments and the value concerns guiding other stakeholder positions. For example, let us return to the public debate regarding whether AquAdvantage should be sold for consumption. If we apply the first three principles (with a human focus), one key concern is highlighted. The salmon is genetically modified to grow to market size in 16 to 18 months, in contrast to the three years that it takes non-modified salmon (Grossman 2016; USDA 2019). Thus, it has the benefit of producing food more quickly for consumption (Principle 2). As the salmon was determined to be safe to eat by the FDA (United States Food and Drug Administration 2019), one could argue that it does not violate the principle of non-maleficence (Principle 3), at least concerning impacts to human health. However, if the company is not required to label their product, decision making capabilities (Principle 1) are not being respected, as consumers have no way of knowing if they are purchasing genetically modified salmon. It should be noted here, however, that the USDA now requires that AquAdvantage salmon be labeled by 2020, as their ruling applies even to products under the jurisdiction of the FDA (Bloch 2019). If this occurs, then the decision-making capabilities of consumers will be respected.

Regarding impacts to society, what are the harms and benefits and how are these distributed? If the price of fresh fish is reduced, as the amount of feed needed to produce the fish was reduced by 25% for the genetically modified salmon (AquaBounty 2019), one could argue that this would provide the company with an economic advantage. This advantage could have negative social impacts, especially for already struggling local producers (Thompson 2006). For example, small-scale and family fishers could be pushed out of business, as sales and the price of fish decline. The harm to fishers, and local communities who rely on the fishing industry, could be seen as an injustice as they are negatively impacted by this technological innovation, while not receiving just compensation for these harms. In addition, it could be seen as a violation of the food sovereignty of communities that rely on salmon or whose history is bound up with the species. Here food sovereignty should be generally understood as a community's right to choose the way that food is produced and consumed (Noll and Murdock 2020). For instance, Virginia Cross, the Muckleshoot Tribal Council Chair argued that "from time immemorial salmon has been central to the culture, religion, and society of Northwest Indian people. Genetically engineered salmon not only threaten our way of life, but could also adversely affect our treaty rights to take fish at our usual and accustomed places" (Northwest Treaty Tribes 2014, n.p.). While the consolidation of fish production may have a plethora of individual benefits, harms to group sovereignty and the marginalization of small-scale producers by mainstream agri-food regimes are pressing areas of concern. Turning to environmental impacts, the FDA has stringent requirements for facilities, including that they be land-based with no access to bodies of water (United States Food and Drug Administration 2019). Based on these requirements, the FDA found that AquAdvantage salmon would not significantly impact the environment. Barring any periphery impacts, if this is the case, then the principles of ecosystem functioning, native species integrity, and maintenance of biodiversity may be satisfied.[2] Finally, concerning animal welfare, the salmon was genetically modified to grow faster but, barring the increased growth-rate, their bodies function in the same manner as non-modified salmon (United States Food and Drug Administration 2019). As such, bodily integrity and species-typical behavior have been maintained. However, quality of life could be an issue, depending on your assessment of fish farming facilities, as the salmon will never experience living in the wild. From this perspective, AquAdvantage salmon largely satisfy four of the five categories of concern. However, the principle of justice is not satisfied, as the livelihoods of fishers and the food sovereignty of Pacific Northwest tribes could be negatively impacted.

The above examination is not meant to be an exhaustive analysis, but is an example of how the GMO Value Framework could be used by individuals and groups to identify substantive issues and guide analysis of future biotechnology applications in the food sector. As stated above, the procedural and relationship components of conflict management are beyond the scope of this paper. However, genomics focused policy decisions, university outreach and communication, STEM education, and agricultural extension conflict management typically include a community engagement component. Leadership and various stakeholder groups engaging in these activities would be better prepared to manage conflict if they have a detailed understanding of the competing values and worries that are most relevant to their genomic application. Using the GMO Value Framework as an analytical tool could help stakeholder groups more easily anticipate potential areas of value conflict concerning individual genomic applications so that they can be used to guide decision making or addressed before tensions are exacerbated and positions are calcified.

For example, AquAdvantage's FDA initial approval in 2015 was challenged due to public concerns that the fish could negatively impact the environment. This approval was upheld years later, but only after a lengthy legal battle. Identifying this substantive concern during the policy process could have enabled the regulatory bodies and/or company to address the concern before public backlash occurred. As Walker (2019) and Ahmadvand and Karami (2007) argue, complex conflicts are guided by competing values and can often never fully be resolved, but they need to be managed so that they do not become intractable or destructive. Identifying the key issues, the sources of value tension, different interpretations of the impact of the technological innovation, and any synergies between groups is a necessary first step for managing and thus unfreezing the current GM debate. This chapter is meant to be a first step in this long process. Future work could focus on strengthening the relationship aspect of conflict situations, as well as building trust between researchers, regulatory bodies, and concerned consumer groups. However, such work is beyond the scope of this chapter. It is my hope that the analytical tools provided herein may move the GMO debate forward, smoothing the way for ethical bioengineering applications in the future.

Notes

1. My analysis uses a condensed version of Stirling and Mayer's (2001) categories of analysis. However, I combined the economic and social categories for clarity, as both capture wider social impacts. In addition, I removed the "ethics" category as this chapter pays particular attention to how normative claims guide arguments that fit into the other categories, as will be discussed in detail in this chapter. Finally, I added animal welfare as a fifth category to Stirling and Mayer's (2001) schema, as impacts to animals play an important role in public discussion concerning GMOs (Thompson 1997; Rollin 2015).

2. It should be noted here that the members of the Pacific Northwest tribe do not agree with this assessment.

References

Ahmadvand, Mostafa, and Ezatollah Karami. 2007. "Sustainable Agriculture: Towards a Conflict Management Based Agricultural Extension." *Journal of Applied Sciences* 7 (24): 3880–3890.

Alkon, Alison. 2014. "Food Justice and the Challenge to Neoliberalism." *Gastronomica* 14 (2): 27–40.

AquaBounty. 2019. "Fast-Growing Genetically Engineered Salmon." *AquaBounty Technologies* (blog). Accessed January 1, 2020. https://aquabounty.com/fast-growing-genetically-engineered-salmon/.

Arneson, Richard. 2013. "Egalitarianism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, n.p. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2013/entries/egalitarianism/.

Barnhill, Anne, Mark Budolfson, and Tyler Doggett. 2016. "Industrial Plant Agriculture." In *Food, Ethics, and Society*, edited by Anne Barnhill, Mark Budolfson, and Tyler Doggett, 407–459. Oxford: Oxford University Press.

Batie, Sandra S., and David Ervin. 1999. "Biotechnology and the Environment: Issues and Linkages." IATP Org. Accessed January 1, 2020. https://www.iatp.org/sites/default/files/Biotechnology_and_the_Environment_Issues_and_L.htm.

Bawa, A. S., and K. R. Anilakumar. 2013. "Genetically Modified Foods: Safety, Risks and Public Concerns: A Review." *Journal of Food Science and Technology* 50 (6): 1035–1046.

Beauchamp, Tom L. 1995. "Principlism and Its Alleged Competitors." *Kennedy Institute of Ethics Journal* 5 (3): 181–198.

Beauchamp, Tom L., and James F. Childress. 2001. *Principles of Biomedical Ethics*. Oxford: Oxford University Press.

Bloch, Sam. 2019. "AquAdvantage, the first GMO Salmon, Is Coming to America." Accessed January 1, 2020. https://thecounter.org/fda-aquabounty-gmo-salmon-seafood-restriction-market/

Byrne, P., D. Pendell, and G. Graff. 2014. "Labeling of Genetically Modified Foods." *Fact Sheet No. 9.371*, Food and Nutrition, 1–4.

Callicott, J. Baird. 1980. "Animal Liberation: A Triangular Affair." *Environmental Ethics* 2 (4): 311–338.

Chambers V, Edgar, Edgar Chambers IV, and Mauricio Castro. 2018. "What Is 'Natural'? Consumer Responses to Selected Ingredients." *Foods* 7 (4): 65.

Christman, John. 2018. "Autonomy in Moral and Political Philosophy." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, n.p. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/autonomy-moral/.

Davison, J., and Y. Bertheau. 2010. "EU Regulations on the Traceability and Detection of GMOs: Difficulties in Interpretation, Implementation, and Compliance." *Cab Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 2 (77): 1–15.

DeMarco, J. P. 2005. "Principlism and Moral Dilemmas: A New Principle." *Journal of Medical Ethics* 31 (2): 101–105.

Doezema, Tess. 2017. "Skepticism About Biotechnology Isn't Anti-Science." Slate. Accessed January 1, 2020. https://slate.com/technology/2017/04/the-aquadvantage-salmon-debate-and-skepticism-about-biotechnology.html.

Losey, J., L. Rayor, and M. Carter. 1999. "Transgenic Pollen Harm Monarch Larvae." *Nature* 399: 214.

Macnaghten, Phil. 2004. "Animals in Their Nature: A Case Study on Public Attitudes to Animals, Genetic Modification and 'Nature.' " *Sociology* 38 (3): 533–551.

Maghari, Behrokh Mohajer, and Ali M Ardekani. 2011. "Genetically Modified Foods and Social Concerns." *Avicenna Journal of Medical Biotechnology* 3 (3): 109–117.

Marchant, Gary E., and Guy A. Cardineau. 2013. "The Labeling Debate in the United States." *GM Crops & Food* 4 (3): 126–134.

Marris, Claire. 2001. "Public Views on GMOs: Deconstructing the Myths." *EMBO Reports* 2 (7): 545–548.

Mather, Damien W., John G. Knight, Andrea Insch, David K. Holdsworth, David F. Ermen, and Tim Breitbarth. 2012. "Social Stigma and Consumer Benefits: Trade-Offs in Adoption of Genetically Modified Foods." *Science Communication* 34 (4): 487–519.

Mellor, David J. 2017. "Operational Details of the Five Domains Model and Its Key Applications to the Assessment and Management of Animal Welfare." *Animals* 7 (8): 60.

Naess, Arne. 1973. "The Shallow and the Deep, Long-range Ecology Movement. A Summary." *Inquiry* 16 (1–4): 95–100.

Noll, Samantha. 2013. "Broiler Chickens and a Critique of the Epistemic Foundations of Animal Modification." *Journal of Agricultural & Environmental Ethics* 26 (1): 273–280.

Noll, Samantha. 2017. "Food Sovereignty in the City: Challenging Historical Barriers to Food Justice." In *Food Justice in US and Global Contexts: Bringing Theory and Practice Together*, edited by Ian Werkheiser and Zach Piso, 95–111. New York: Springer Publishing.

Noll, Samantha, and E. Murdock. 2020. "Whose Justice Is It Anyway? Mitigating the Tensions Between Food Security and Food Sovereignty." *Journal of Agricultural & Environmental Ethics* 33 (1): 1–14.

Northwest Treaty Tribes. 2014. "Muckleshoot, ATNI, Oppose Genetically Modified Salmon." Northwest Treaty Tribes. Accessed July 10, 2020. https://nwtreatytribes.org/muckleshoot-atni-oppose-genetically-modified-salmon/.

Owen, L., W. Howard, and M. Waldron. 2000. "Conflict over Farming Practices in Canada: The Role of Interactive Conflict Resolution Approaches." *Journal of Rural Studies* 16: 475–483.

Palmer, Clare. 2010. *Animal Ethics in Context*. New York: Columbia University Press.

Phillips, Theresa. 2008. "Genetically Modified Organisms (GMOs): Transgenic Crops and Recombinant DNA Technology." *Nature Education* 1 (1): 213.

Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge: Harvard University Press.

Regan, Tom. 1995. "Obligations to Animals Are Based on Rights." *Journal of Agricultural & Environmental Ethics* 8: 171–180.

Rich, Matthew. n.d. "The Debate over Genetically Modified Crops in the United States: Reassessment of Notions of Harm, Difference, and Choice." *Case Western Reserve Law Review* 54 (3): 889–916.

Rollin, Bernard. 2012. "The Moral Status of Invasive Animal Research." In *Animal Research Ethics: Evolving Views and Practices*, edited by Susan Gilbert, Gregory E. Kaebnick, and Thomas Murray, S4–S6. Garrison, NY: The Hastings Center.

Rollin, Bernard. 2015. *The Frankenstein Syndrome: Ethical and Social Issues in the Genetic Engineering of Animals*. New York: Cambridge University Press.

Rollin, Bernard E. 2006. "The Regulation of Animal Research and the Emergence of Animal Ethics: A Conceptual History." *Theoretical Medicine and Bioethics* 27 (4): 285–304.

Ronald, Pamela. 2016. "The Truth about GMOs." In *Food, Ethics, and Society*, edited by Anne Barnhill and Tyler Doggett, 485–502. Oxford: Oxford University Press.

Sandler, Ronald. 2005. "Book Review: Gregory Pence, Editor, The Ethics of Food: A Reader for the 21st Century." *Journal of Agricultural & Environmental Ethics* 18: 85–93.

Savulescu, Julian. 2011. "Genetically Modified Animals: Should There Be Limits to Engineering the Animal Kingdom?" In *The Oxford Handbook of Animal Ethics*, edited by Tom Beauchamp and R. G. Frey, 641–671. New York: Oxford University Press.

Shrader-Frechette, Kristin. 2002. *Environmental Justice: Creating Equality, Reclaiming Democracy*. Oxford: Oxford University Press.

Singer, Peter. 2015. *Animal Liberation: The Definitive Classic of the Animal Movement*. Open Road Media.

Stirling, Andy, and Sue Mayer. 2001. "A Novel Approach to the Appraisal of Technological Risk: A Multicriteria Mapping Study of a Genetically Modified Crop." *Environment and Planning C: Government and Policy* 19 (4): 529–555.

Tester, Mark. 2001. "The Dangerously Polarized Debate on Genetic Modification." *British Food Journal* 103 (11): 785–790.

Thompson, Paul. 1993. "Ethical Issues Facing the Food Industry." *Journal of Food Distribution Research* 24 (1): 12–22.

Thompson, Paul. 1997. "Ethics and the Genetic Engineering of Food Animals." *Journal of Agricultural and Environmental Ethics* 10: 1–23.

Thompson, Paul B. 2006. "Should We Have GM Crops?" *Santa Clara Journal of International Law* 4 (1): 75–95.

Toft, Kristian. 2012. "GMOs and Global Justice: Applying Global Justice Theory to the Case of Genetically Modified Crops and Food." *Journal of Agricultural and Environmental Ethics* 25 (2): 223–237.

United States Food and Drug Administration. 2019. "AquAdvantage Salmon Fact Sheet." Accessed January 1, 2020. http://www.fda.gov/animal-veterinary/animals-intentional-genomic-alterations/aquadvantage-salmon-fact-sheet.

Walker, G., and S. E. Daniels. 1997. "Foundations of Natural Resource Conflict: Conflict Theory and Public Policy." In *Conflict Management and Public Participation in Land Management*, edited by B. Solberg and S. Miinna. Joensuu, 13–36. Finland: European Forest Institute.

Walker, Gregg. 2019. "Assessing Collaborative and Transformative Potential via the 'Progress Triangle:' A Framework for Understanding and Managing Conflicts." Oregonstate.Edu. Accessed January 1, 2020. http://oregonstate.edu/instruct/comm440-540/triangle.htm.

Walker, Rebecca. 2006. "Human and Animal Subjects in Research: The Moral Significance of Respect versus Welfare." *Theoretical Medicine and Bioethics* 27: 305–331.

Waltz, Emily. 2017. "First Genetically Engineered Salmon Sold in Canada." *Scientific American*. Accessed January 1, 2020. https://www.scientificamerican.com/article/first-genetically-engineered-salmon-sold-in-canada/.

Warren, Karen. 2000. *Ecofeminist Philosophy: A Western Perspective on What It Is and Why It Matters*. New York: Rowman & Littlefield.

Webster, John. 2016. "Animal Welfare: Freedoms, Dominions, and 'a Life Worth Living.'" *Animals* 6 (6): 35.

Zhao, Y., H. Deng, C. Yu, et al. 2019. "The Chinese Public's Awareness and Attitudes toward Genetically Modified Foods with Different Labeling." *Science of Food* 3 (17): npj.

CHAPTER 26

THE MINDED BODY IN TECHNOLOGY AND DISABILITY

ASHLEY SHEW

1. Introduction

I often wonder exactly when I became a cyborg. (Am I truly one yet? I wonder that too.) After all, the technologies implanted in me are not the techno-magical devices of science fiction, aimed at transcending the human body's limits. The devices implanted in me are mostly there to stop things from happening that I did not want to happen.

The IUD (intra-uterine device) installed by my midwife (very painfully, I might add) kept my menstrual periods from worsening and kept me from having another pregnancy. The port-a-cath that was installed on my chest—under my skin, around my collar bone, and into a large vein—kept the strong chemotherapies that were required for my bone cancer from burning up the peripheral veins in my arms. I could have gotten this removed after my chemotherapy was over, but I didn't. It has been over five years, and I do not think I will ever get it removed. The cancer has recurred twice, although no chemo was required for either recurrence and only surgical intervention, but it feels better just to have the port at the ready. I go to the cancer clinic every eight weeks so they can flush liquids through it to keep it working.

Although the terrible chemotherapy made my IUD unnecessary by rendering me infertile, I kept that in me, too. It is an artifact of my body's history. At some point, it will have to be extracted. The rods and screws in the leg that doctors had to amputate helped force my bones together and make my leg secure, made my amputated leg more functional. They kept me from needing a more traditional amputation, and they help me avoid phantom pain. The prosthesis I wear on this leg comes off. *Does it count as one of my cyborg parts?* It's not *in* me, but the more I wear it the more it feels like part of me, of my experience in getting around in the world—until it doesn't any more, and I need a new one. My leg keeps me from falling. I think about Laurie Anderson's lyrics: "You're walking. And you don't always realize it/But you're always falling" (Laurie Anderson, "Walking and Falling," *Big Science*, 1982). My leg helps keep me from falling further, keeps me from the ground, but my transition to this cyborg stance was not seamless. I re-learned how to walk, learned how to wear this leg, and the fake leg often reminds me that it is not my original one. The thinginess of things (to lift a phrase from philosopher Davis Baird) is not to be underestimated. I should not romanticize this heavy piece of metal and carbon fiber and Velcro that is both part of me and not.

If the prosthesis counts as part of cyborg me, what about my walker? I love the smooth movement of my walker rolling across a wood floor, and it feels like part of me sometimes too. We flow together. I take off this seemingly heavy object, which is what my prosthesis feels like at the end of a long day, and glide through my living room. *Can parts that detach and interchange still be considered part of my extended body?*

And what about my hearing aids? They, too, detach. But unlike my other cyborg parts, they don't seem to be fending off unwanted happenings. They let me hear the shy student in the back row of my class when she has a question. But I still ache to take them off by the end of the day, and feel relief in their removal. They keep me hearing, but they often keep me hearing things—air conditioners, washers, whooshes, ringing tones, beeps, and creaks—that would otherwise not be focal in my experience.

Poet Jillian Weise, whose leg is computerized, has written of people she calls tryborgs—tech enthusiasts who wish they were (or declare that they are) cyborgs, yet fail to recognize the true cyborgs among them, technologized disabled people. These tryborgs, and Weise names names, consist of transhumanist thinkers like Ray Kurzweil, Michio Kaku, and many others. As Weise puts it, "The tryborg tries to integrate with technology through the latest product or innovation." In addition to "big names" in transhumanism, she talks of those who are first adopters on things like Google Glass and other gadgets, imagining themselves taking up a brave new future, ahead of the curve, when it comes to new technologies. Transhumanists (those who identify themselves with the movement to enhance the human condition through the pursuit and adoption of technologies) and other tryborgs imagine a day where machine and human merge into one and see themselves as paving this avenue. Weise writes:

> Tryborgs rely on the nonexistence of actual cyborgs for their bread and butter. If cyborgs exist, how will the tryborg remain relevant? Wouldn't we just ask the cyborg for her opinion? The opinions of cyborgs are conspicuously absent from the expert panels, the tech leadership conferences and the advisory boards. The erasure is not news to us. We have been deleted for centuries, and in the movies, you will often see us go on a long, fruitful journey, only to delete ourselves in the end.
>
> (Weise 2016)

When I look around the cancer clinic, I think to myself, *Wow, a lot of cyborgs out today*. Sick people, chronically ill people, disabled people—we are what cyborgs look like, but this is far from the popular image of the cyborg.

Indeed, our cyborg status is rarely recognized. The techno-sublime visions of the future provided by popular imaginings in science fiction and transhumanist scholarship rarely feature authentic disabled people. We get the Million Dollar Man and the Terminator, with near-magical powers, and we have the technological imaginations of transhumanists dreaming of the elimination of disability and the delay of death. Ray Kurzweil, whose original work was in reading machines for the blind during the 1970s, has argued for the elimination of death and for the re-creation of people after their deaths through computing and the merging of human consciousness with machine. He seeks pills and other life extension so that he can make it to the Singularity—a sudden acceleration of growth in computational power that will enable biological bodies to become "irrelevant," because we can upload human consciousness into supercomputers (Kurzweil 1999).

Transhumanist (and presidential candidate for the political Transhumanist party in the past few US elections) Zoltan Istvan anticipates a world where exoskeletons are funded and sidewalks can be left in disrepair (Istvan 2015). It's unclear whether he's spoken with anyone disabled about their experiences with inaccessibility. He writes:

> As the 2016 US Presidential candidate of the Transhumanist Party, I advocate for doing whatever is necessary to eliminate physical disability altogether. We are shortchanging our citizens and our country by not doing otherwise. In the 21st Century, with so much technology and radical medicine at our fingertips, we should reconsider the Americans with Disability Act. It's great to have a law that protects against discrimination, but in the transhumanist age we also need a law that insists on eliminating disability via technology and modern medicine.
>
> (Istvan 2015)

Many disability journalists and advocates have written against this position—for the logical flaws in his analysis on choosing the false dilemma of exoskeletons or sidewalks, the fact that we already have the tech for sidewalks and people are disabled now, and, in fact, some disabilities can be exacerbated by standing. The fact is that many wheelchair users do not want exoskeletons—or would only use them for limited tasks, not as a replacement for wheelchair technologies (Eveleth 2015; Sauder 2015; Ladau 2015; Peace 2007–2019). Many wheelchair users see their wheelchairs as a source of their freedom in the community, not a problem. Disabled people often adapt to technologies at hand—and are not always, as usually depicted, in a perpetual state of wanting "better" technology where better means making people seem more nondisabled. So much in television and movies and even popular news stories indicates that disabled people are inadequate as disabled—when they show such characters and people in stories at all.

When science fiction movies do show disabled characters, they are almost never actually played by disabled people. Instead, science fiction shows us supposed futures in space that leave disabled people behind, either through cure or passive elimination. The reality (as cyborgs tell us if we listen) is that disabled bodies are uniquely suited for technological futures in space, where we will all become impaired in some way (Shew 2017). Disabled bodies can do more than is ever depicted; they already operate in ways that consider spatial awareness more carefully (Wong and de Leve 2017), and disabled people have an uncanny knack for hacking the environments and tools around us (Jackson 2018).

In this chapter, I first discuss the narrative accounts and tropes that exist around cyborg bodies, exoskeletons, prostheses, and other wearable tech—tech that is often aimed at mediating or "fixing" disability. I then describe traditional accounts of body and mind, and how these traditional accounts have shaped the ways that cyborg bodies are imagined. But I also think that these traditional accounts are wrong, in some ways, and in the later part of this essay, I offer a critique of traditional accounts of body and mind, grounded, in part, in recent work in disability studies. This critique more importantly draws from the stories of the cyborgs among us, the disabled people who already experience technological "enhancement." They (we) know what it is like to live with bodies and minds significantly altered by technology—or what it is like to live with bodies and minds that people anticipate will be "improved" by technology. I conclude this chapter with words of resistance to techno-optimistic narratives about bodies and minds that serve to flatten the experiences of real cyborgs.

2. Accounts of Cyborg Bodies, Exoskeletons, and Prostheses

> Maybe tryborgs imagine that the theorist Donna Haraway's "A Cyborg Manifesto" is right. The manifesto reads: "In short, we are cyborgs. The cyborg is our ontology; it gives us our politics." But Haraway is a tryborg: she's not disabled; she has no interface; she uses the term as a metaphor. The strategic move where one group says, "I shall speak for them because they do not exist / do not live here / do not have thoughts" is common of the tryborg. When they are not speaking for us, they may take a detour into animal studies, a field where they can rest assured that their subjects remain silent.
>
> (Weise 2016)

Conversations about "enhanced" bodies are dominated by the voices of tryborgs, transhumanists, and other techno-enthusiasts, and nearly all media coverage of prosthetic and exoskeletal technologies emphasizes the "life-changing power" of these devices. Media coverage features interviews with the makers (almost never the users) of these technologies, but there is rarely any mention of cost, ease of transition, availability, weight of device, and user abilities—all issues that factor into whether or not these technologies are available to or even usable by the people the technology is aimed at. Nobody asks what users actually want (Eveleth 2015). Instead, media stories are suffused with romantic ideas about the restoration of ability, wholeness of body, and shiny new futures represented by these technological interventions.

In this transhumanist future, human beings and machines merge, and all humans are "enhanced." Sometimes these futures are imagined in terms of replacing body parts with more useful ones—often, for some reason, increasingly complicated (and sometimes weaponized) prosthetic arms. Elon Musk argues for his technology Neuralink, a high-bandwidth brain-machine interface where neural electrode threads assist in merging human brain with machine (Musk and Neuralink 2019). Sometimes these futures are imagined in terms of exoskeletons, which promise to enhance strength and endurance and to enable fragile human bodies to go into hostile environs. And sometimes these futures involve the complete replacement of human bodies, with our minds downloaded into mechanical, synth-bio forms or plugged into a virtual world.

Transhumanist bioethicist James J. Hughes, known in disability circles for his appearance in the documentary *Fixed: The Science/Fiction of Human Enhancement* (2013), tells us that: "The horror and enthusiasm that our cyborg future excites clearly have more to do with the transgression of the body's boundaries than with the actual enhancements it will bring, since those enhancements are or will be accessible far more cheaply, safely and upgradably in wearables and gadgets" (Hughes 2014, 26). This dream/assumption of increased accessibility, cheapness, and safety often rides on successful testing, an existing market, and adoption. The dream of affordability is often not met from the perspective of many disabled people, especially when we think about costly prosthetics and fancy drugs under patents. I am wearing about $15,000 of leg and hearing aids as I write this, and my body is relatively inexpensive compared to many disabled people I know; this figure does not include the cost of my port-a-cath, daily medications, any other surgical "upgrades" to my body, nor replacement and maintenance costs. Being promised the latest and greatest often puts people under pressure to perform as a good techno-optimistic disabled person—entering yourself as a test pilot (or, worse, guinea pig). Disabled people regularly have their body's boundaries "transgressed"—willingly and with good reason because we like to live. That's not the hang-up here.

For transhumanists, assistive technologies for disability seem to be pilot projects aimed at an assumed future when everyone will use such technology, and disabled folks are simply test cases for these technologies: "feeling" prostheses that are connected to amputees' nerves, exoskeletons to be tested by wheelchair users, and computers that interface with the brains of stroke patients, those with traumatic brain injuries, and paralyzed patients to enable them to play games and move mechanized parts. By using disabled people to test wearable tech, engineering for transhumanist futures is framed as engineering in a humanitarian mode.

Many engineers and designers can't even comprehend that what they want to make is not always what disabled people want to use. I once sat down with a student "Design for America" group that was focused on 3D-printed hands. We spent well over an hour talking about why people choose *not* to use prosthetic hands, even when they are available. Some people do not want better prosthetic arms, and some do not want prosthetic arms at all—a fact that surprised all the students involved. The students had never imagined that someone simply might not want a device that was available and cheap—the aim of their project. They thought cost was the only factor. I used evidence from memoir, blogs, and narratives written by arm amputees (some of whom I know and others to whom I am grateful for sharing publicly) to talk with them about the myriad factors that go into decisions about prosthetic arms and hands. We talked about the pressure to wear prostheses, to bring disabled bodies as close to "normal" as possible: amputees who don't wear prostheses (or who choose older technologies) are publicly questioned by nondisabled people, who (primed by transhumanist-inflected media for enthusiasm about prostheses) ask about their choices, finances, and even whether it is appropriate for them to be seen in public without prostheses that make others feel comfortable.

Similar things happen to wheelchair users as the public is increasingly bombarded with news and whiz-bang rhetoric about exoskeletons. Not all wheelchair users would be candidates for exoskeletons; many wheelchair users can already stand or walk short distances, but standing causes seizures or drops in blood pressure. But many people look to exoskeletons to "fix" the "problem" of using a wheelchair. For many wheelchair users, their wheelchair use is not a problem to be solved. The problem is mostly a matter of designed spaces: they need better sidewalks and curb cuts and working elevators. The social model, as theorized in disability studies, frames disability in this way. This model and others are often theorized against the backdrop of the medical model of disability that says that disability is an individual medical impairment (Wasserman, Asch, Blustein, and Putnam 2016; Barnes 2016). The experiences of individual disabled people and the historical development of the category of disability speak against a thin and under-nuanced medical understanding of disability. The excitement about exoskeletons from the nondisabled public and tech innovators outstrips the excitement from the disabled people for whom the technology is ostensibly being created.[1]

In fact, the technologies and "hacks" that disability communities get most enthusiastic about (usually technologies that don't have revolutionary or exciting applications for nondisabled people) are ignored by the press in favor of sexy innovations like exos and fancy prostheses. For example, can we talk about how great and liberatory the redesign of walkers was? Now that wheeled walkers have a basket to carry some stuff and a platform on which to sit, I can carry my laptop across the room all by myself. (Now, designers, please just add cup-holders, but not ones that will scratch against the sides of doorways please.)

3. Traditional Accounts of Body and Mind

> [U]nderstanding intelligence and the body in the way I've described suggests that AI researchers ought to be thinking not only about *how* intelligent creatures are intelligent, but also w*hy* they are intelligent . . . The body is what produces this need, what anchors intelligent creatures in the world, what *invests* us in it, what makes the world relevant and significant to us, what makes it such that we *have to cope*. Bodies, in a word, are why intelligence matters.
>
> (Susser 2013 286)

There is no way to cover the whole history of philosophical reflection about the nature of our minds and how they relate to our bodies. In this section, I offer a sketch of a few major ways in which this relationship is conceived, focusing primarily on accounts that have been extensively referenced and have influenced how the general public thinks about their own minds and bodies.

The paradox of the Ship of Theseus has been discussed by philosophers from Plutarch to Thomas Hobbes to Daniel Dennett. This paradox asks us to consider how entities are related to their parts: in this thought experiment, a ship is built for Theseus, and over the course of its history, boards are replaced, with the ship eventually no longer having a single remaining original board. Is this still the Ship of Theseus? Hobbes ([1655] 1839) asks a further question: What if a new ship were made, board by board, out of the pieces from the original Ship of Theseus? Would that be the Ship of Theseus? Which ship then "counts" as the Ship of Theseus: the original form with entirely new boards, or the entirely new form made out of all the original pieces?

This paradox is not only about ships. As our cells die off and are replaced, do we remain the same people we started out as? How do we understand identity—of ships, of people—when parts are replaced over time, whether they be shipboards, cells, or, now, body parts? The relationship of identity to bodily structure—the question of what counts as an original or a true identity in the face of inevitable bodily change—is grounded in our view of the relationship between mind and body.

One notion of this relationship that has had an enduring hold on philosophy is Rene Descartes' mind-body dualism, expressed most clearly in his *Meditations* (Descartes [1641] 1998). Mind-body dualism is the idea that the mind and the body are completely distinct and separate entities (or different "substances," which is why this view is sometimes called *substance dualism*). According to Descartes, mind and body can exist separately from one another, and they are united only by the intervention of a deity. In this theory, the mind and body intersect in a specific place in the body, the pineal gland of the brain, but our thinking-stuff—our mind—remains separate from the body. Although many philosophers push against this idea, the foundational binary—the notion that there are two types of stuff, thinking-stuff and body-stuff—is difficult to resist. Indeed, this binary of stuff is reinforced, again and again, by each argument against it, as people argue that the two (separate stuffs) are connected and contingent. This binary distinction, these two concepts ideated as separable even in how we speak about their union, shapes what we can think about.

Even as David Hume argues against Descartes, offering us a different way of thinking about the relationship between mind and body, he retains the foundational idea that there is a binary (a mind and a body) in human interactions with the world (Hume [1896] 2011). For Hume, our minds and bodies are more closely related than for Descartes. Our minds are shaped by the sensory data we take in through our bodies; there is no mind that exists before sensory experience. Ideas are built up from sensory experiences in the world, not abstractions poured into our minds by a deity. We learn to make judgments and inferences as the mind reflects on experience, and experience only comes with a body. Thus, for Hume, mind and body are in constant interaction—mind forms with bodily, sensory experience after birth; the body and associated sense experience is pre-rational, prior to organization, and the mind learns to make sense of gathered experiences.[2] Hume, in arguing against Descartes, still relies on a binary between bodies and minds, which he sees as two separate things that, in interaction, make sense of the world together.

More recent work into the relationship of mind to body is driven by research into artificial intelligence, machine learning, and how the environment shapes perceptions. Daniel Dennett has criticized the inherent separability posited in mind-body dualism, arguing that, for example, we cannot know under dualism whether people are zombies (bodies without souls or minds)—including whether we ourselves are zombies (Dennett 1991). According to Dennett, what we think of as mind arises out of the physical: bodies produce minds, without reference to souls or deities. Dennett sees consciousness as a gradualist, evolutionary phenomenon, a sliding scale. Though he has his detractors (who often think there is more to consciousness than the current scientific explanation or that consciousness doesn't give the whole picture), Dennett's ideas have influenced a growing community of scholars studying consciousness and mind, often scholars of AI and animals.

Philosophers such as Andy Clark and David Chalmers argue that the body and perception are heavily intertwined with consciousness (1998). Robert D. Rupert neatly summarizes their views: "Human perceptual processing makes information about the environment available to cognitive mechanisms that implement a wide range of cognitive functions, from scientific reasoning to the physical navigation of the immediate environment" (Rupert 2010, 4). This position is known as the extended mind thesis, for it posits that cognitive processing is exclusive neither to the body nor to the mind. Rather (contra Dennett), cognition takes place in interactions between body, brain, and environment. They support an "active externalism" view, holding that the environment actively drives and produces cognition. Clark and Chalmers write, "beliefs can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes" (1998, 12). This coincides well with Mark Rowlands' take on the extended mind thesis, arguing that the environment itself is part of the extended mind. He calls this theory of radical externalism, which he developed, in part, out of findings from animal studies and studies of animal cognition, "environmental epistemology." And, as Rowlands notes, "The environment that an organism can manipulate or exploit includes not just inanimate structures but also other creatures" (Rowlands 2005, 11). He thus radically expands the location and complexity of cognitive processes. For Rowlands, the entire environment becomes a model of cognition; in his view, we are not only individual bodies or minds, for we are deeply engaged with the world in our experience and cognition.

These last views offer some resistance to the mind-body separability, but still work within and against frameworks that set up the dualism. Donna Haraway's positioning of the cyborg against the dualism of multiple binaries (discussed next) is similar to Andy Clark's move to eliminate another binary through claiming that we are all already cyborgs (Clark 2003). In both cases, Jillian Weise would warn us against tryborgs and against those who would imagine we didn't exist in order to create us, or to create themselves as such.

4. The Imagined Cyborg Body Results from Traditional Accounts

The idea arising from the history of philosophy of mind—with Plato, Descartes, and Locke—that our bodies and minds are separable entails that minds are downloadable, transferable, discrete—that we can move those minds to other bodies. This concept, though much older, is reflected in our science fiction. *Star Trek*'s transporters present a literal Ship of Theseus scenario; *Star Trek: The Next Generation's* Data's cognition arises out of his physical properties, and we see the same notion in the reboot of *Westworld*, in *Altered Carbon*, *Travelers, Dollhouse*, *Person of Interest, Ghost in the Shell*, and *Blade Runner*, where we see literal downloads or transfers of minds into various different bodies.[3]

Many transhumanists predict and imagine that we will someday be able to download ourselves or parts of ourselves into tech devices (Kurzweil 2013; Fourtané 2018). To believe this, one must believe that our minds are *things* that can be detached from their contexts, that they are part of a mind/body binary. This is the notion that grounds philosophical ideas about cyborgs, which purport to go beyond binaries but instead simply invert them or reinforce them. This is so even outside the transhumanist literature. For example, Donna Haraway's formulation of the cyborg was rooted in an attempt to get beyond the binary of patriarchy and its response, eco-feminism (Haraway 1991). Haraway tried to find a way out of this binary by taking a political stance as a cyborg, which is neither simply organic or technologic alone. This philosophical move revealed the ironies involved in commitment to a binary, but, as Jillian Weise has noted, produced another ironic binary: unlike Weise, who has computerized body parts, Haraway uses the cyborg as a metaphor only, and is not herself a literal cyborg.

Just as Haraway wanted to escape the 1980s binary between traditional patriarchy and eco-feminism (and the hierarchy that framed patriarchy as the more privileged of the two positions), we want to be able to escape the dualism of mind-and-body. But like Haraway's manifesto for cyborgs, this attempt often simply replicates the original binary while inverting the hierarchy. When Dennett (1991), for example, argues that body constitutes mind, he does so by making a distinction between body and mind, replicating the binary even as he tries to get past it.

This distinction between body and mind continues to weigh heavily in all cultural discussions of bodied technologies. If we could just get the right shape of mind, we feel, we could produce technology that exploits or captures that mind and re-body it into another form. When we imagine we can exist as just-minds in something else—flesh or metal—we imagine that it is possible to separate mind from body. When we imagine we can merge with tech, it is because we see technologies as upgrades to what we have, but this means we ignore the reality of maintaining, debugging, and carrying with us real (not imaginary, perfect, science-fiction) technologies as constant companions.

Add to this considerations around disability: Victoria Pitts-Taylor, placing philosophy about the brain and body in the context of minds and in relation to critical disability studies, argues that "bodily morphology and the environment are not separate elements with independent epistemic contributions. Rather, ability and disability—and other kids of difference as well—can be seen in terms of the different ways body-minds couple or fit with various other elements in the world . . . rather than seeing disability as an essential category of the body-subject" (Pitts-Taylor 2016, 57).

Given a selective reading of human history, one can think the progression to cyborg is inevitable—humans can come to look and be like "natural-born cyborgs" (Clark 2003). But in fact, the imagined cyborg of the future grows directly out of the philosophical tradition that sees mind and body as different things. In contrast, recent work by Rowlands, Pitts-Taylor, and others suggests an environmental or social-cultural embeddedness that makes the notion of a simple transition of minds and selves into technology much harder to imagine.

5. The Existing Cyborg Body

Are there better ways to think about cyborg tech? If philosophy has (mostly) gotten it wrong, where should we look for better, more grounded, less binaristic thinking about bodies and minds, technology and humanity? I argue that tools for thinking about minds, bodies, and environments are best captured by recent work in disability studies, as Jillian Weise suggests when she reminds readers that cyborgs already exist.

According to Daniel Susser (2013), in a reading of Hubert Dreyfus and Mark Bickhard (two current theorists who consider embodiment as fundamental to intelligence), "The more or less discrete physical systems we call bodies are just the sort of physical systems with the capacity to interact skillfully with their environments. The distinction between bodies and intelligence is an analytical distinction—it refers to two aspects of the same phenomenon (its physical properties and its skills or capacities)" (285). He argues that Dreyfus sees "that intelligence and the body are inseparable, that they are two sides of the same coin, that they develop together in the world, that intelligent creatures are intelligent because they are embodied" (258).

This idea—that body and intelligence (minds) are inextricably one—coincides with a recent turn in disability studies initiated by Margaret Price's borrowing of the word *bodymind* from trauma studies. She writes, "because mental and physical processes not only affect each other but also give rise to each other—that is, because they tend to act as one, even though they are conventionally understood as two—it makes more sense to refer to them together, in a single term" (Price 2015, 269). Bodyminds are one whole—this is true for disabled and nondisabled people; our experiences are never extricable in the experience of one or the other. Even outside considerations about disability, the evidence is piling up that body and mind are deeply connected, as shown by the gut-brain interaction, the physical toll of mental health, the mental toll of physical exclusion, and the phenomenon of phantom pain.

The notion of bodyminds captures the way that any truly cyborg tech would and could be assimilated. We aren't body and mind. We're always already both, and our experiences reflect this. Modern-day cyborgs—people with disabilities and people who are chronically ill—know this better than anyone. As our "cyborg laureate" Jillian Weise puts it, "I know it will take time, but things will change. For a while, all the experts on African-Americans were white. All the experts on lesbians were Richard von Krafft-Ebing. All the experts on cyborgs were noninterfaced humans" (2016). If we are interested in cyborgs—in creating them, in being them—there is much to learn from "interfaced humans"—the disabled and other technologized folks (often people who are chronically ill) among us.

Postphenomenologists in philosophy of technology have also recognized the imbrication of bodymind and of technology and the body. They emphasize the mediation technologies provide in our thinking and our actions. Postphenomenologists work on the phenomenon of phantom vibration (Rosenberger 2015), on multistability (Wellner 2015), and other experiential accounts of technologies. Scholars like Don Ihde, the above-quoted Daniel Susser, Peter-Paul Verbeek, and Stacy Irwin, among others, examine how technologies are situated in our experience, what becomes focal in a technology's use and what remains obscure, flexibility and creativity in our uses of technology, and the experiential changes that happen to individuals in response to technological change. This research trajectory comingles mind and body as well as environment. As phenomenologists have long talked about lifeworlds, so it makes sense to talk about *the lifeworlds of bodyminds*—our experiences are always contingent on our environments and our bodyminds (this united word to encapsulate the us, the thing that does the sensing, reacting, and processing of the cacophony the world presents). Lifeworlds indicate that our experiences of this world are never mere ideas or recorded data; they are concretely lived perspectives and relationships to the world and everything else, and always already mediated by technology. Adding to this the concept of bodymind points to a sort of unity of experience rarely captured in Western philosophical writing.

Vivian Sobchack (once a student of postphenomenologist Don Ihde) has described how the technology of crutches profoundly changes the experience of moving through the world:

> If one learns how to use crutches properly, they are extraordinarily liberating. Indeed, one can move more quickly and with greater exuberance on crutches than on one's own two legs (whether prosthetic or not). The span of one's gait increases and there is a cadenced and graceful "swing through" effect that not only covers ground but also propels the lived body forward in pleasingly groundless ways not allowed by mere walking. There is, both phenomenologically and empirically, a "lift" to one's step.
>
> (Sobchack 2005, 59)

When we listen to what disabled people (rather than tech designers) write and say about technology, we find a resituation of expertise about technology, one that values the lived experiences of bodyminds. Technologies are valued differently, with less flashy technologies valued much more highly. We attend to mundane technologies and their maintenance—things that designers caught up in the techno-exuberance of our current era often neglect (Russell and Vinsel 2016) to the detriment of what disabled people actually need and want (Earle 2019).

Cassandra Crawford, in her book *Phantom Limb* (2014), has suggested that listening to disabled people as cyborgs is crucial for reasons beyond the practical feedback they can offer on designing usable cyborg technology. For Crawford, disabled people and cyborgs are both inherently destabilizing and revolutionary:

> amputees inhabit a unique position relative to our understanding of disability and embodiment because cyborgs decenter able-bodiedness. Even as pure fantasy, they represent what becomes of those who join "the revolution" and as such, they prefigure the future of lived techno-corporeal conjoin-ment.
>
> (Crawford 2014, 248)

But all disabled people, not only amputees, can do this work in people's imaginations. Although popular imagination about amputees paints us as inspirational, heroic, brave people, glamorized by the technologies that seem to work so well,[4] this comes as part of the spread of tropes about who *deserves* these technologies, about how disabled people's lives will be dramatically transformed (read: brought into line with "normality") by those technologies, and about how engineers creating assistive tech are humanitarian innovators who should be praised. All this coverage ignores the real problems that make disabled people's lives difficult in the first place: poor infrastructure and planning, disregard of disability rights activists' messages about what the community needs, the cost of being disabled (both in cost of technologies and in social and time costs), and ableism in the systems that disabled folks must navigate. And often this coverage neglects less apparent disabilities and issues of pain and its management (something for which people do hope for better solutions).

6. The Politics of Cyborg Expertise

Technological design practice ought to affirm disability as a valuable category of identity and experience and appreciate disabled bodyminds while looking to address issues important to those in the community. Questions of access, intentions versus impact, cultural understanding and expertise, and cultural narratives that bend expectations become critical in this sphere. In this sphere, new movements in disability activism and disability studies elevate cyborg[5] ways of knowing as design expertise.

Feminist Standpoint Theory (Harding 2003) resonates some of this emphasis on situatedness and perspective, drawing from the work of philosophers like Sandra Harding, Helen Longino, Allison Jaggar, Nancy Hartsock, and others. The standpoint of cyborgs matters if we want to talk about what it is like to be technologized humans operating in various environments and with varieties of technological interventions. To put it simply: disabled cyborgs have important knowledge to lend when it comes to technologies and infrastructure, knowledge that has been shaped and tempered by close relation, failure, and being. Often, disabled people would prefer we affirm disability as an identity category through how we design, rather than design that aims to make someone nondisabled (or easier to closet or pass).[6]

When disabled people declare cyborg expertise, this is a political claim, because cyborg voices have for so long been ignored and dismissed. Too few disabled people have been consulted about our cyborg-ness, our technological choices, or the planning and design (usually done by non-disabled people) that shapes our lives. Although not all disabled people will be cyborgs, technologies are often seen as the answer to the question "What shall we do about disability?" But when people ask that question, what they really mean is "What can we do about disabled people?" This evokes a long history of neglect and exclusion, of questions and policing (Samuels 2015),[7] of institutionalization, eugenics, sterilization, abuse, lobotomies, and electric shocks[8]—of trying to "fix" disability or even of thinking that disability is what unworthy people deserve, so they deserve to suffer (read: either we want to cure you, or you brought this on yourself, for which you do not deserve lives equal to those of nondisabled people). All of these things are, of course, exacerbated by bias and exclusion in our technosocial and political systems, which are set up by non-disabled people to serve non-disabled people. The experiences and treatment of disabled people are often informed by their race, class, gender, and other factors (Thompson 2016, Barbarin 2016).

Many voices in disability studies and disability community activism are calling for disabled people's expertise to be recognized. This expertise is essential and important, but almost completely absent from crucial conversations on infrastructure, planning, social welfare, and even disability issues themselves. Disabled people are the objects rather than the authors of social policy. To attempt to rectify this, the disability civil rights movement in the United States has taken as its rallying cry "Nothing About Us Without Us," asserting that decisions about disability ought to include disabled people (and the respect for the authority of disabled people on matters of disability).

Since even before the advent of the independent living movement in the 1970s, disabled people have been uniting and organizing to create and preserve independence for community members. That work is ongoing, and recent political action by ADAPT has brought disability rights to greater public prominence. ADAPT, originally founded to push for public transportation that is fully inclusive of disabled people, has been a powerful force in other political actions that affect disability communities: saving social health programs like Medicaid, fighting for continued home-based care, and pushing to end the torture of institutionalized people that still continues today (ADAPT, n.d.).

As disability activists have asserted disability expertise on political social policy that impacts the community, disability scholars have begun to study disabled people's expertise in design. Liz Jackson (2018) has written powerfully about design by disabled folks, and more broadly about disabled people as "the original lifehackers." She has also lobbied publicly against an exhibit on disability and design curated by nondisabled people; curators and designers themselves often take disabled people as unpaid, unattributed consultants. This is not uncommon in the disability world: we are often asked to provide our opinion on things without attribution, recognition, or compensation. This is why we must keep declaring our particular expertise, and why our voices should be recognized.

Aimi Hamraie and Kelly Fritsch offer perhaps the most direct call to action on the topic of technology and design with "The Crip Technoscience Manifesto" (Hamraie and Fritsch 2017; Hamraie and Fritsch 2019), which even before its 2019 publication already had the community abuzz from its presentation at conferences. Echoing the title of Donna Haraway's famed Manifesto for Cyborgs, also like Haraway in occupying a space in feminist approaches to considering how people are cast and understood, "The Crip Technoscience Manifesto" offers a different way to imagine and think about bodies and technologies. This manifesto anchors a special issue on Crip Technoscience of *Catalyst*. Hamraie and Fritsch explain in the original 2017 version:

> Broadly, crip technoscience challenges the presumption that valuable scientific knowing and technological change proceed from neutral, non-disabled bodyminds. Instead, we offer crip technoscience as a project premised upon interdependence, desiring disability, critical design, and user-expertise. (2017)

In its originally presented form, our authors describe seven hallmarks of crip technoscience:

1. "Crip technoscience is a polemic against imperatives for cure and normalization."
2. "Crip technoscience centers affiliation and interdependence."
3. "Crip technoscience aspires toward accessible futures."
4. "Crip technoscience elevates disabled ways of knowing as design expertise."
5. "Crip technoscience is activist technoscience, contested and politicized knowing-making."
6. "Crip technoscience marks design as a 'desiring practice'—a way of affirming disability as a desirable onto-epistemological practice of being and relating."
7. "Crip technoscience agitates against empire, compulsory normalcy, militarization, and mandates for productivity." (2017)

Hamraie and Fritsch give us guideposts toward understanding disability-centered design within an inclusive, situated, dependent frame that takes seriously disabled people as experts and disability as a rich category of identity and political affiliation. Our authors see disabled cyborg situatedness: "Disabled people, Alison Kafer argues, often have 'an ambivalent relationship to technology' " (2017). Cyborgs, in the dreams of tryborgs, are techno-enthusiasts who adopt and test and wear their tech—all while desiring more synthesis with machines. Actual lived experiences of technologized humans—disabled and chronically ill people—are far more complicated and less wildly optimistic. Technologies are offered to disabled people as miracles, but rarely are they felt that way, and sometimes we resist: as in the well-known case of cochlear implants in Deaf culture, in the current work of folks positioned against exoskeletons, and in opposition to the soaring prosthetic narratives offered in our culture that rarely map onto the experience of amputees.

Cyborgs, that is, cyborg bodyminds, share stories different from what the dominant cultural narrative and transhumanist rhetoric appreciate or allow. It is also worth noting here that notions of time and space and experience are changed by disability and technology. Concepts like Crip Time and Deaf Time within disability and Deaf studies indicate as much (Bauman and Murray 2014; Samuels 2017). The environments in which we operate (the worlds in which we experience things) are fundamentally changed by shifting categories of identity, embodiment, and technology.

7. Conclusions that Should Upset Traditional Notions and Whiz-bang Rhetoric

Alison Kafer (2013) writes about the use of the word cyborg in feminist and disabled communities, describing the signs and resistance offered by Laura Hershey and Connie Panzarino. Hershey, a former poster child for the Muscular Dystrophy Association, led protests against the organization and its host Jerry Lewis, who served to further stigmatize her disability and infantilize and make pitiful those with it. Connie Panzarino, a grassroots disability activist and out lesbian who used a tracheomotomy, proudly displayed during Pride marches the sign “Trached dykes eat pussy without coming up for air” (which Kafer discusses, 122). Kafer writes:

> In common parlance, Hershey and Panzarino could be considered “severely disabled” (Haraway’s “severely handicapped”). They rely on power wheelchairs; they employ personal attendants to assist them in their daily activities; and their chronic impairments occasionally led to medical crises, particularly respiratory ones. For most cyborg theorists, the story would stop there, serving as a perfect illustration of the ways in which (certain) bodies don’t end at the skin. Indeed, in this framework, the more severely disabled one is, the more cyborgian, because the more likely to be using high-tech medical equipment and adaptive technologies. A crippled cyborg politics, however, refuses to stop with this kind of recitation of diagnosis or condition. Following Robert McRuer, “severe” can be read as defiance, fierceness, critique . . . Rather than reduce these activists’ experiences to the details of their impairment, let us focus instead on their complex and contradictory negotiations with

> technology, or on the ways in which such negotiations lead to questions about community, responsibility, pleasure, and complicity. (2013, 124)

In this article, I've tried to suggest how philosophical ideas about the separation of bodies and minds, and their imagined future relations to technology, have overlooked consideration of real cyborgs—disabled crips—and our expertise. Our ideas and experiences are always within context and connected deeply to our bodyminds. That any groups speak of a coming cyborg future—as if cyborgs don't yet exist (or only few do)—without the input of one of our most technologized groups is the height of erasure. Perhaps what such thinkers will find from the disability community, with experiences that defy exciting predictions and resist simplification into an optimistic narrative, is not what they wish to hear and cannot appreciate. Ambiguity does not bring all the Venture Capitalists to the yard. Indeed, crip politics tends to lean away from such promotionalism and appreciate that, even when good gadgets are made, affording and benefiting from them can still be difficult, given the system we live in. Context always matters.

However, if we, as philosophers of technology, seek narratives that reflect relationality and embeddedness, defy simplification, and stay true to human experience, disability politics and crip technoscience can help guide our thinking about how humanity interfaces with technology, the trade-offs in our technological futures, and the cultures made and habits reinforced by our technological choices.

Notes

[1] William Peace's blog, Bad Cripple, offers one source of testimony to this claim. He has numerous posts on exoskeletons.

[2] This has an interesting parallel to how a person learns to use a prosthesis, I think. The feedback one gets from one's body as one becomes accustomed to a new device is not normalized in experience yet, and it takes time and use before one's sensory load from a device comes to make sense. While the Humean picture of the mind in body allows for our ability to grow accustomed or habituated to particular technologies, Hume is right that it takes time, and habituation is often harder and less pleasant than generally recognized.

[3] Thanks to my colleagues Damien Williams and Kristen Koopman for their expertise in forming this list.

[4] This is very different from the actual experiences of many amputees, however. The media images are mostly of athletic, youthful amputees who lost their leg through trauma, often military service related, but, as Crawford notes, statistically, the "typical prosthetized amputee is likely to be male, below-knee (BK amputee), who is older, African American, and has lost his leg to vascular disease" (2014, 248).

[5] Disability scholar Bethany Stevens even coins the word "cripborg" to emphasize disability—and disability in its proudness and badassery with the word "crip," derived from cripple and being reclaimed by some disability activists and scholars (Nelson, Shew, and Stevens 2019).

[6] See also the work of Elizabeth Barnes in *The Minority Body* (2016) for discussion of a theory of disability as identity, rather than impairment.

7. Krip Hop Nation, Sins Invalid, ADAPT, and the Autistic Self-Advocacy Network, among other groups, have brought visibility to these issues too.
8. Which, as of 2019, still happens at the Judge Rotenberg Center, and which has been the center of action by activists with ADAPT. #StopTheShock

References

ADAPT. n.d. "ADAPT—Free Our People." Accessed July 11, 2020. http://adapt.org/

Barbarin, Imani. 2016. "Things I've Learned in This Black Disabled Female Body." *Crutches and Spice*, July 8. https://crutchesandspice.com/2016/07/08/things-ive-learned-in-this-disabled-black-female-body/

Barnes, Elizabeth. 2016. *The Minority Body: A Theory of Disability*. Oxford, U.K.: Oxford University Press.

Bauman, H-Dirksen L., and Joseph J. Murray, eds. 2014. *Deaf Gain: Raising the Stakes for Human Diversity*. Minneapolis: University of Minnesota Press.

Clark, Andy. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. 2003. Oxford, U.K.: Oxford University Press.

Clark, Andy, and David J. Chalmers. 1998. "The Extended Mind." *Analysis* 58, no. 1: 7–19. http://www.jstor.org/stable/3328150

Crawford, Cassandra S. 2014. *Phantom Limb: Amputation, Embodiment, and Prosthetic Technology*. New York, NY: New York University Press.

Dennett, Daniel. 1991. *Consciousness Explained*. New York, NY: Back Bay Books.

Descartes, René. [1637, 1641] 1998. *Discourse on Method and Meditations on First Philosophy*. 4th ed. Translated by Donald A. Cress. Indianapolis: Hackett Publishing Company.

Earle, Joshua. 2019. "Cyborg Maintenance: Design, Breakdown, and Inclusion." In *Design, User Experience, and Usability. Design Philosophy and Theory*, edited by Aaron Marcus and Wentao Wang. HCII. Lecture Notes in Computer Science, vol. 11583. Springer.

Eveleth, Rose. 2015. "The Exoskeleton's Hidden Burden." *The Atlantic*, August 7. https://www.theatlantic.com/technology/archive/2015/08/exoskeletons-disability-assistive-technology/400667/

Fixed: The Science/Fiction of Human Enhancement. 2013. DVD. Directed by Regan Brashear. Newburgh: New Day Films.

Fourtané, Susan. 2018. "Neuralink: How the Human Brain will Download Directly from a Computer." *Interesting Engineering*, Sept. 2. https://interestingengineering.com/neuralink-how-the-human-brain-will-download-directly-from-a-computer

Hamraie, Aimi, and Kelly Fritsch. 2017. "Crip Technoscience Manifesto." Conference presentation, Society for the Social Studies of Science Annual Meeting, Boston, August 30–September 2.

Hamraie, Aimi, and Kelly Fritsch. 2019. "Crip Technoscience Manifesto." Catalyst 5, no. 1: 1–34.

Haraway, Donna. 1991. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." in *Simians, Cyborgs, and Women: The Reinvention of Nature*. London, U.K.: Free Association Books.

Harding, Sandra, ed. 2003. *The Feminist Standpoint Theory Reader*, 1st edition. Routledge.

Hobbes, Thomas. [1655] 1839. "Of Identity and Difference," in *The English Works of Thomas Hobbes of Malmesbury*, edited by Sir William Molesworth. London: J. Bohn. https://archive.org/details/englishworksth021hobbgoog

Hughes, James J. 2014. "How Conscience Apps and Caring Computers Will Illuminate and Strengthen Human Morality." *In Intelligence Unbound: The Future of Uploaded and Machine Minds*, edited by Russell Blackford and Damien Broderick, 26–34.West Sussex, UK: Blackwell/John Wiley and Sons.

Hume, David. [1896] 2011. *A Treatise of Human Nature*. Oxford, U.K.: Clarendon Press.

Istvan, Zoltan. 2015. "In the Transhumanist Age, We Should Be Repairing Disabilities, Not Sidewalks." *Motherboard*, April 3. https://www.vice.com/en/article/4x3pdm/in-the-transhumanist-age-we-should-be-repairing-disabilities-not-sidewalks

Jackson, Liz. 2018. "We Are the Original Lifehackers." *New York Times*, May 30. https://www.nytimes.com/2018/05/30/opinion/disability-design-lifehacks.html

Kafer, Alison. 2013. *Feminist, Queer, Crip*. Bloomington, IN: Indiana University Press.

Kurzweil, Ray. 1999. *The Age of Spiritual Machines*. Viking Press.

Kurzweil, Ray. 2013. "Ray Kurzweil: Your Brain in the Cloud." *Big Think, 19 Feb*. https://www.youtube.com/watch?v=0iTq0FLDII4

Ladau, Emily. 2015 "Fix Discriminatory Attitudes and Broken Sidewalks, Not Humans." *Motherboard*, April 8. https://www.vice.com/en_us/article/d73947/fix-discriminatory-attitudes-and-broken-sidewalks-not-humans

Musk, Elon, and Neuralink. 2019. "An Integrated Brain-Machine Interface Platform with Thousands of Channels." bioRxiv, August 2. https://www.biorxiv.org/content/10.1101/703801v4.full.pdf+html

Nelson, Mallory Kay, Ashley Shew, and Bethany Stevens. 2019. "Transmobility: Possibilities in Cyborg (Cripborg) Bodies." *Catalyst* 5, no. 1: 1–20.

Peace, William J. 2007–2019. *The Bad Cripple Blog*. http://badcripple.blogspot.com/

Pitts-Taylor, Victoria. 2016. *The Brain's Body: Neuroscience and Corporeal Politics*. Durham, NC: Duke University Press.

Price, Margaret. 2015. "The Bodymind Problem and the Possibilities of Pain." *Hypatia* 30, no. 1 (November): 268–284. doi:10.1111/hypa.12127

Rosenberger, Robert. 2015. "An Experiential Account of Phantom Vibration Syndrome." *Computers in Human Behavior* 52 (November): 124–131. doi:10.1016/j.chb.2015.04.065

Rowlands, Mark. 2005. "Environmental Epistemology." *Ethics and the Environment* 10, no. 2 (Autumn): 5–27. http://www.jstor.org/stable/40339102.

Rupert, Robert D. 2010. *Cognitive Systems and the Extended Mind*. Oxford, U.K.: Oxford University Press.

Russell, Andrew, and Lee Vinsel. 2016. "Hail the Maintainers." Aeon.co, April. http://aeon.co/essays/innovation-is-overvalued-maintenance-often-matters-more

Samuels, Ellen. 2015. "Will the Real Disabled Person Please Stand Up, or What's Wrong with This Picture." Invited Talk at Virginia Tech, Blacksburg, VA, September 25.

Samuels, Ellen. 2017. "Six Ways of Looking at Crip Time." Disability Studies Quarterly 37, no. 3. https://dsq-sds.org/article/view/5824/4684

Sauder, Kim. 2015. "When Celebrating Accessible Technology Reinforces Ableism." CrippledScholar, July 4. https://crippledscholar.com/2015/07/04/when-celebrating-accessible-technology-is-just-reinforcing-ableism/

Shew, Ashley. 2017. "Technoableism, Cyborg Bodies, and Mars." Technology and Disability, November 11. https://techanddisability.com/2017/11/11/technoableism-cyborg-bodies-and-mars/

Sobchack, Vivian. 2005. "Choreography for One, Two, and Three Legs." *Topoi* 24, Issue 1 (January): 55–66. doi:10.1007/s11245-004-4161-y

Susser, Daniel. 2013. “Artificial Intelligence and the Body: Dreyfus, Bickhard, and the Future of AI.” In *Philosophy and Theory of Artificial Intelligence*, edited by Vincent Müller, 277–287. Berlin: Springer-Verlag Berlin Heidelberg.

Thompson, Vilissa. 2016. “Black Disabled Woman Syllabus: A Compilation.” Ramp Your Voice, May 5. http://www.rampyourvoice.com/black-disabled-woman-syllabus-compilation/

Wasserman, David, Adrienne Asch, Jeffrey Blustein, and Daniel Putnam. 2016. “Disability: Definitions, Models, Experience.” *Stanford Encyclopedia of Philosophy*, May 23. https://plato.stanford.edu/entries/disability/

Weise, Jillian. 2016. “Dawn of the Tryborg.” *New York Times*, November 30. https://www.nytimes.com/2016/11/30/opinion/the-dawn-of-the-tryborg.html?_r=0

Wellner, Galit. 2015. *A Postphenomenological Inquiry of Cell Phones: Genealogies, Meanings, and Becoming*. Lexington Books.

Wong, Alice, and Sam de Leve, eds. 2017. Crips in Space Issue. *The Deaf Poets Society* 4 (May). https://www.deafpoetssociety.com/issue-4/

CHAPTER 27

OUTER SPACE AS A NEW FRONTIER FOR TECHNOLOGY ETHICS

KEITH ABNEY

1. Introduction

Space poses a complex set of problems, to which our geocentric intuitions about how to proceed with technological development are liable to be as ill-adapted as our bodies are to weightlessness. Geologists aver the Anthropocene Era has begun on Earth; that is, humanity's impact will be discernible through geological time. Similarly lasting impacts will soon alter our corner of the Solar System, beyond a lonely American flag on the Moon. So, how should we be responsible stewards of space—what marks do we want to leave on the neighborhood?

All too often, even highly educated humans tend to approach novel situations based on their experience in other realms; all too frequently, that "inside the box" thinking misleads us into suffering unforeseen, devastating harms. We should seek to avoid unacceptable risks in a new environment *before* discovering them the hard way, by making some initial group suffer. This essay investigates this "*first-generation problem*" (Abney 2017a) for space; what risks can we reasonably foresee when we discard terrestrial assumptions?

After all, space represents a true frontier: immense distances from society with remote prospects of a safe return if things go wrong, few human pioneers as of yet, extreme environments, and great physical hardships. And frontiers are always easier to navigate if you have a good map of the dangers beforehand.

2. Economic and Social Concerns

2.1 First, Questions of Moral Status

A crucial issue in space ethics is moral status: what kind of value should non-human life or technology have? Rejecting speciesism, that is, a morally unjustifiable value-preference for one's own species, does not solve the issue. Nonhuman alien persons, like the Star Trek characters Worf or Spock, would clearly deserve moral consideration. But what is the moral value of alien bacteria, or even fishlike creatures, in the oceans of Europa? Numerous options on the scope of intrinsic moral value have been defended (Lupisella 2010), including all life (biocentrism), or, indeed, everything in the universe (cosmocentrism). Other common views include rationality/sapience as crucial (ratiocentrism; e.g., Abney 2004, 2018), or tying moral value to sentience ("sentientism"; e.g., Singer 1974)—the capacity for conscious experience of pleasure and pain.

Regardless of the correct account of what has intrinsic value, it is a mistake to conflate having intrinsic value (which by definition requires no extrinsic, external relationship) and final value (valuable as an end in itself, not as a mere means to some further end). Christine Korsgaard (1983) usefully distinguishes these two types of moral value. A key realization for technology ethics is that intrinsic value requires final value, but the converse is false. Assuming that the existence of morality requires beings with moral responsibility (Abney 2019), then only morally responsible agents have intrinsic value; but many other things could have final value. Hence, all other moral value depends on what agents *should* value—whether it is purely instrumental (like an asteroid valued for its mineral wealth), or a final value, like an aesthetic pleasure in merely viewing Saturn's rings, which may exist solely in the mind of a valuer (and so not be intrinsic), but also is not a mere means to any further end.

Despite ongoing controversy on the topic of moral status, one conclusion should be clear: the use of robots and other pieces of technology (as long as they lack moral agency) has a practical moral superiority over using humans for almost all our near-term legitimate purposes in space. As long as robots have no intrinsic moral status (for more detail, see Abney 2012; Veruggio and Abney 2012), there is no intrinsic wrong in using them as a means to explore, mine, and even help humans colonize alien locales. (Of course, their use may still violate certain important final values, and so be unethical in a different way.)

Further, suppose we encounter alien persons, or have as a final value unspoiled alien landscapes, flora, or fauna; it's best if the first representatives of humanity to encounter them are creatures which are morally dispensable—our robotic scouts and ambassadors, our proxies and advocates for human civilization. It may turn out to be very important indeed to make a good first impression, to be willing to sacrifice our creations for our would-be friends; as opposed to making them our enemies. In fact, this willingness could become a matter of existential importance; an argument to be continued in the final section.

2.2 Space Economics and the Ethics of Risk

Scenario 1: Imagine an asteroid mining company (like Planetary Resources) that has taken possession of a small asteroid. After removing the valuable minerals for transshipment to Earth, Saudi Arabia offers them $100 million if they will fashion a streamlined kinetic projectile packing a nuclear bomb-level punch (but without the radiation), otherwise known as a "rod from God" (Stilwell 2018) from the slag and launch it towards a concentration of Houthi fighters in Yemen, or their allied base in Iran. Is there any way to ensure this will not happen, or to stop it if it does? Does a relevant legal entity exist to enforce any such restriction, or would something like a "space court" have to be invented? Would the answers change if a private organization or business (e.g., al Qaeda, Facebook, Rosneft) attempted something similar?

This first scenario highlights the intersection of economic, environmental, political, and military concerns about space and technology that make it a focus of dual-use worries. The 1967 Outer Space Treaty (OST) is the preeminent international legal basis for understanding many of the issues here, but on many issues it is worryingly vague. On issues from property rights to extraterrestrial objects, current interpretations of the OST are problematic, given its language about the "provenance of all mankind" and the Article II claim that "Outer space, including the Moon and other celestial bodies, is not subject to national appropriation by claim of sovereignty, by means of use or occupation, or by any other means." These seemingly imply private property cannot exist in space. But where some see a prohibition, others see a loophole; it is not clear under international law if Article II also applies to private actors, as independent companies like SpaceX and Blue Origin were not envisioned by the drafters in 1967. Could private citizens own property in space, even if no nation-state can claim sovereignty over such property? (Would that make, e.g., taxing space property impossible?)

Current United States law (Obama 2015) and the actions of private companies like Planetary Resources (2018) indicate a belief in such space property. So, which is it? International law here is woefully underdeveloped for the emerging technologies. What model should emerging space law and policy use? Should the guiding analogy for space development be Antarctica (Skibba 2018), the Wild West, a new territory of an existing nation-state, or something else?

Many economic purposes for space development appear *prima facie* morally defensible, because they pose no immediate, direct, and grave terrestrial risk, but do promise benefits to (at least some of) humanity. They include: developing and exploiting energy resources, finding/creating new sources of food and water, developing and exploiting mineral resources through mining, engaging in space-based manufacturing and construction, establishing space colonies and other space settlement, and the use of space for defense and security concerns, from mere surveillance all the way to engaging in war from space. But ethical consideration of such technologies will not be exhausted by their (lack of) direct harm to Earth-bound humans.

2.3 Ethical Issues for Mining and Construction

One common moral argument for space-based technological development in manufacturing, mining, and construction uses cost-benefit analysis. Astronomer and science popularizer Phil Plait writes: "Some estimates say that for every dollar invested in the Apollo program, more than 20 have been returned. That's a huge payoff! Computer tech, communications, rocketry, and many other fields have benefited hugely from space exploration." (Plait 2007) But the costs may be underestimated. Like many industries, the space industry in low-Earth orbit threatens to have a serious pollution problem, in this case with space debris. One terrible possibility is the Kessler syndrome (1978), dramatized in the movie *Gravity* (2013), in which pieces of debris begin colliding and breaking apart, causing yet more collisions, until a runaway cascade renders entire orbital slots unusable. The UN report "Space Debris Mitigation Guidelines of the Committee on the Peaceful Uses of Outer Space" (2010) details the threat to the long-term sustainability of activities in low-Earth orbit, and investigates how to mitigate those dangers. Any serious cost-benefit analysis must consider the longer term, to ensure low-Earth orbits remain usable for future generations.

The problem is a version of the tragedy of the commons (Hardin 1968): no one currently owns low-Earth orbits, but anyone who can get satellites there benefits. Hence, the incentives are to further crowd the area until it risks becoming unusable. Paradoxically, increasing space debris only exacerbates the incentive problem: if your satellite may malfunction because it will run into space junk (and thereby itself become additional space junk), then your private selfish incentive is to put up additional satellites as a backup; when everyone thinks this way, the problem rapidly escalates in severity. So, given the apparent "market failure," should the responsibility to clean up space junk lie with private contractors, nation-states, NGOs, the UN, or some new organization?

Scenarios that involve a "tragedy of the commons" are most commonly solved either by privatization—taking the commonly held good and partitioning it into slices of private ownership (such as selling orbital slots)—or by top-down regulation (such as making and enforcing rules on all private space-faring entities.) Either solution would require an international organization to create and *enforce* (realistically, using only robots!) a treaty that would oversee near-Earth orbits in a more rigorous way than the extant Outer Space Treaty.

Recognizing this problem, the EU and others have proposed solutions in "The International Code of Conduct for Outer Space Activities" (Space Code 2014). Such documents are testimony that space may require wholesale revision of the usual economic and social models, which assume terrestrial regulations, rewards, and punishments—and a human presence. Current ideological approaches to the novel challenges of the final frontier seem unlikely to produce a happy outcome; it will require sophisticated ethical and policy analysis.

2.4 Ethical Issues for Manufacturing

There will also be specialized manufacturing in space. The key driver will be "additive manufacturing," better known as 3D printing. It has already begun: 3D printers began manufacturing on the ISS in 2014 (Snyder 2014). Additive manufacturing may revolutionize space missions; there's an ESA project to design a lunar base using 3D-printed moon rock. Even sooner, the zero-G vacuum of space allows for industrial processes impossible on Earth, such as creating a hollow titanium lattice ball with a complex internal geometry (Chao 2014).

Once asteroid and other space mining becomes routine, logistics could be revolutionized in space, and on Earth (Snyder 2014). Raw materials could be taken from elsewhere in the solar system and processed cheaply in Earth orbit (and beyond) compared to the expense of bringing such materials to space from the Earth's surface. Products could be made available on demand using "just in time" manufacturing with easy delivery all over the world—just wait until the desired drop point in orbit is reached, and down it goes. Much of the land-based infrastructure for shipping could become obsolete; far fewer cargo ships, delivery trucks, or train shipments may be needed.

Further, manufacturing toxic or other risky products could be done off-world: potentially hazardous waste or otherwise dangerous aspects of manufacturing could be quarantined in space without risk to the environment of the Earth (or human settlements). Already, there are proposals to dispose of nuclear waste by firing it into the Sun (Cain 2017). The combination of zero-G and plentiful raw materials also allows the production of megastructures; things too big to construct on the surface of the Earth could be assembled in orbit for use in space, or even (carefully!) returned to the Earth's surface for use there.

Space manufacturing will not require humans—except perhaps as passengers. If we find (or create) another habitable world and want to travel there *en masse*, the construction of a "space ark" requires automated space manufacturing (Nielsen 2014). There are serious moral questions as to whether we should build ships that intend to send many generations of humans to live and die in an attempt to reach the stars (Levy 2016), but in all likelihood, if such starships ever come to exist, they will be built by robots in space.

2.5 Energy and Environmental Issues

Many private companies plan on introducing advanced technology into space; their ethical theorizing rarely goes beyond adherence to law and possible risk to humans. But if alien ecosystems or even alien landscapes have final value, such approaches may portend wildly unethical exploitation of hitherto pristine frontiers. Hence, environmental ethics intersects outer space technology concerning both planetary protection of space-originating life (Persson 2008; Schneider 2013) and many varieties of space development, even those of lifeless worlds (Lupisella and Logsdon 1997; Arnould 2009). Space technology could also affect terrestrial environmentalism. Space-based solar power (SBSP) could give us clean energy while simultaneously reversing global warming, without befouling the atmosphere with aerosol pollutants (Dorminey 2017), as other geoengineering proposals routinely plan.

One proposal (Mankins 2014) involves large autonomous robotic solar arrays in orbit that could collect SBSP and beam it in microwave form back to Earth; these robots will either have the capacity for self-repair and automatic solar-orientation to maximize collected energy, or have additional robots tasked for those jobs. In Japan, Mitsubishi and JAXA have already built proof of concept wireless transmitters for solar power from space, and both NASA and Japan have a goal of having robotic built space-based solar power beaming from geostationary satellites by the late 2030s (Rodriguez 2013). JAXA has a further goal of having robots build a 12 mile-wide, 6,800 mile-long “Luna Ring” of solar panels to be constructed on the moon’s surface. This massive belt would beam solar power straight to Earth via microwaves and lasers (Singh 2011).

There are several advantages to SBSP over traditional solar. First, the uninterrupted nature of the sunlight collected in space; no need to worry about cloudy days or precipitation blocking sunlight. Second, SBSP requires very little real estate to implement, and can therefore be efficiently stationed directly over population centers. This avoids the transmission losses from having extensive panels in the middle of deserts with output delivered hundreds of miles over power lines (Mankins 2014). Third, if global cooling through solar radiation mitigation is an additional goal, the panels could be made much larger and designed to maximize the amount of sunlight blocked from hitting the Earth, and stationed over areas most in need of cooling—say, the North and South Poles. Fourth, space-based solar panels don’t have to worry about bird poop or dust or any other atmospheric pollutants, and so require minimal maintenance.

Space could also provide another source of abundant energy beyond solar, one that may be key to solving both Earthly concerns for power generation and the propulsion needs of future spacecraft. The technology is nuclear fusion. But fusion needs fuel, specifically deuterium and helium-3—and there is a bounty of both in the atmospheres of Jupiter, Saturn, Uranus, and Neptune. As Bryan Palaszewski (2014) writes:

> Atmospheric mining in the outer solar system has been investigated as a means of fuel production for high energy propulsion and power. Fusion fuels such as Helium 3 (3He) and hydrogen can be wrested from the atmospheres of Uranus and Neptune and either returned to Earth or used in-situ for energy production.

So, space resources and technology may provide a radical “green” solution to our energy crises. But that requires answering a key question: who should own and profit from space resources—individual persons, private corporations, nation-states, NGOs, or no one?

2.6 Owning Property in Space

Despite (or because of?) the uncertainty in interpreting the OST, President Obama (2015) signed a bill recognizing asteroid ownership rights by United States citizens, and encouraging their exploration for commercial use. Given the legal go-ahead, numerous groups (e.g., Deep Space Industries, Kleos Space, Planetary Resources, etc.) plan to engage in asteroid mining and other extraction of extra-planetary resources (Cornish 2017). But even if legal, this potential new "gold rush" raises thorny issues. Regulators will need to determine how to weigh the economic value of such activities versus aesthetic, ecological, epistemic, and other values, up to even human health and survival. Are such values commensurable with economic values? Policymakers need to know the correct approach to take: cost-benefit analysis, risk-benefit analysis, or something else. All these questions require a sophisticated ethical analysis to be creditably answered, lest space potentially devolve into a Hobbesian war of all against all.

Space offers valuable resources beyond mining minerals on asteroids. Is water on asteroids (Schwartz 2016), or at the Moon's south pole, or at the Martian ice caps, a freely available good, or should there be strict regulations on harvesting it (Boyle 2011)? The Moon Treaty (1979) was supposed to begin to settle such questions, but no spacefaring nation has yet signed it. What about resources that have clear dual-use capabilities as both fuel and weapons (e.g., mining an asteroid for uranium, or an atmosphere for deuterium or helium-3)? These are fundamentally philosophical questions, as our current conceptions of property emerged out of debates among thinkers like Locke and Rousseau. We need to revisit the terrestrial assumptions that these thinkers labored under to develop a robust and novel theory of property in space (Simberg 2012).

To see this need for new thinking, consider the following: can robots own property? One tradition in political philosophy going back at least to Locke and exemplified more recently by Robert Nozick (1974) sees the justification for the ownership of private property as derived from either free trade of property one already rightfully owns, in exchange for the rightfully owned property of another; or, for the initial acquisition of property, one mixes one's own labor with a freely available good, as long as there remains "good enough and plenty" raw materials left over for others likewise to engage in such original acquisition. For the bounty of potential goods in space, on this view, the only question is who will claim them first, by performing the labor of getting to them and justly acquiring the resource for themselves. Taken seriously, such a view seems to imply that we will need the labor of humans in space in order to legitimate claims to private property there, at least initially. Unless, that is, a robot (perhaps one representing Planetary Resources or Virgin Galactic or Kleos Space) can claim ownership! Further problematizing the issue is that the OST does not easily comport with a Nozickian understanding of justly acquiring private property. Instead, it expressly delimits who and what may use space resources, without endorsing a "first finders, keepers" approach.

3. Space-Based Dual-Use Technology

Space-enabled technologies like SBSP or nuclear fusion using 3He could have immensely positive environmental and social consequences. But there are negatives as well; in addition to tremendous start-up costs, nuclear fusion or large-scale SBSP is clearly dual-use, capable of fulfilling both civilian and military roles. Choosing which parts of Earth's surface to shade could easily become an act of aggression, even war (Abney 2017a). So how should dual-use technologies be regulated, and who gets to decide?

3.1 Armed Conflict in (or from) Space

Military discussions of space often refer to "the ultimate high ground" (Posey 2014). But what does that even mean? If the claim is that it provides the ideal platform for launching WMDs (weapons of mass destruction), that appears doubly false: first, orbital weapons systems are difficult to make stealthy, and so would make easily identified, predictable targets. Second, the OST explicitly prohibits WMDs. But, what constitutes a clear WMD/non-WMD distinction in space? If only nuclear weapons count, then the OST seemingly allows many other highly destructive possibilities, as basic physics means a kinetic weapon (like the "rod from God" in Scenario 1) that causes only moderate damage when launched from Earth, could cause immense destruction when descending from orbit.

Further, if potentially dual-use space equipment primarily serves a clear civilian purpose, then a state that preemptively attacks it would likely start a major conflict; one the other side could reasonably argue was unprovoked. So, policymakers who want to avoid conflict desperately need to clearly demarcate purely civilian from dual-use from purely military purposes in space technology. Alas, that seems impossible: vessels used for space exploration and other research, manufacturing and construction, or even rescue could also be used in military operations. Spaceports and search-and-rescue stations could be used as military bases. As mentioned, geoengineering systems like SBSP could be used as warfare devices, by shading key areas of Earth. Abuse of seemingly reasonable uses of space remains all too likely, with potentially lethal unintended consequences. Space tourism offers an instructive example.

3.2 Space Tourism and Dual Use

Space tourism could be a cover for a dual-use program. Indeed, the Pentagon reportedly planned for the same space planes that take civilians joyriding to also transport UAVs or even human troops to a distant battlefield quickly:

> Roosevelt Lafontant had a dream. A Marine Corps officer assigned to the National Reconnaissance Office at the Pentagon, Lafontant had a back-row seat in late 2001 as the Marines spearheaded the invasion of landlocked Afghanistan. To reach

> Kandahar from their assault ships, the Marines had to fly more than 400 miles over Pakistan in rickety, heavy-lift helicopters. "There's got to be a better way," Lafontant recalled thinking.
>
> (Axe 2014)

Space tourism is often appealed to as a potentially legitimate reason for a human presence in space. But the military has viewed space tourism as a pretext for developing the capacity to rapidly deploy troops from near-orbit to remote, isolated battlefields. In theory, SUSTAIN (an acronym for "Small Unit Space Transport and Insertion" (Axe 2014)) could deploy forces from the United States to anywhere in the world within two hours. Flying at sub-orbital altitudes, SUSTAIN theoretically would be invulnerable to enemy air defenses, and it could avoid violating the national airspace of countries bordering the war zone.

SUSTAIN was supposed to be incognito: disguised as part of a venture for lifting tourists into space, like Virgin Galactic. The dual-use would be simple: to change from space tourism to war, simply switch out the passengers and retarget the coordinates. But (officially, at least) SUSTAIN is not being developed, so shock troops descending from space is on hold outside of movies like *Starship Troopers*. However, the United States does have the X-37 robotic space plane. It stays aloft for weeks to months, and officially carries no weapons, but if it did, it could launch "rods from God" or any other weapons system that would fit in its hold almost anywhere in the world in a manner of hours. The Russians are trying to build a similar space robot with the capacity to fire nuclear weapons anywhere around the world within 2 hours–an ability they believe the X-37 may already have (Axe 2016).

Nonetheless, even the X-37 is not destabilizing in the way the SUSTAIN program would be; having human or robotic troops ready to swoop down at any moment raises fundamentally different strategic concerns from those associated with mere (sub-)orbital flying robots. Generally, dual-use concerns are exacerbated by a human presence in space. After all, the civilian astronauts or tourists may secretly be spies or soldiers preparing an attack from the ultimate high ground. In addition to personally causing an attack, humans may accomplish nefarious ends by stealth: they may be able to override whatever safety measures are in place, either by an in-person cyberattack, or even by physically overriding or destroying security features of spacecraft. Humans could also engage in other kinds of subterfuge undetectable from the ground; they could reorient satellites, or change their orbit to encounter debris, and so on. These concerns would be alleviated somewhat by only having robots in space; civilian robots could still be hacked and repurposed for military attacks, but short of that, dual-use problems with civilian spacecraft are minimized when no humans, only robots, are allowed into space.

3.3 The Space Force and the History of Militarizing Space

Even before President Trump (2018) announced plans for the Space Force as a sixth branch of the US military, prominent politicians had already floated plans for clearly military or at least dual-use purposes in space. For instance, in his 2012 presidential campaign, Newt Gingrich wanted the United States to have a manned base on the Moon by 2020. Later NASA feasibility studies endorsed the idea, and the Constellation program had it as one of its goals (Whittington 2015). Dual-use issues were clear in Gingrich's pitch, despite any military activities on a lunar base explicitly violating the Outer Space Treaty. He broached the idea that the colony would be the "51st state" and made clear the potential military aspects of such a colony when he said "We will have commercial near-Earth activities that include science, tourism, and manufacturing, and are designed to create a robust industry precisely on the model of the development of the airlines in the 1930s, because it is in our interest to acquire so much experience in space that we clearly have a capacity that the Chinese and the Russians will never come anywhere close to matching" (Gingrich 2012). Arguably, the only way a colony could attract the necessary funding is by having a military purpose: not even counting the initial costs of getting there and construction, Phil Plait (2012) estimated simply maintaining even a small colony would take at least $7.4 billion per year, over 1/3 of NASA's budget.

Gingrich and others could make such plans without much blowback, because the OST has loopholes that have permitted weapons testing in space by the United States and Russia since the 1960s (Union of Concerned Scientists 2012). Dual-use was the key: anti-satellite weapons (ASATs) can perform military strikes on enemy spacecraft, but are permitted by the OST because they have the civilian justification of deorbiting derelict satellites. By the 2000s, more countries and new technologies were involved. The United States and others developed satellites that can maneuver and approach targets, as well as laser systems designed to interfere with satellite sensors. In 2007, China used a mobile, ground-based missile to launch a homing vehicle that destroyed an aging weather satellite by "kinetic kill," resulting in a record level of persistent debris and engendering real worries about the Kessler syndrome (Johnson-Freese 2009). Perhaps in response, in 2008 the United States demonstrated the ASAT capabilities of its sea-based Aegis missile defense interceptors by destroying a non-responsive satellite at an altitude of 240 km. Many other nations are now acquiring the capability for war in space: for example, in 2010, India also announced its intentions to develop a kinetic kill ASAT system (Union of Concerned Scientists 2012). North Korea now has missiles capable of intercepting satellites as well.

This history of dual-use concerns and technology development, all permitted by the letter of the Outer Space Treaty, provokes questions about the future regulations for the use of space. There are numerous suggestions as to how the OST should be renegotiated in the light of emerging dual-use technology, including how to make it an ongoing process (regulations will need further updating as technical capabilities change). Many nations are cognizant of the need for such updates. In 2008, Russia and China presented a draft treaty—the Treaty on the Prevention of the Placement of Weapons in Outer Space, known as the PAROS treaty (2017). While it would limit the use of ASAT weapons, it would do nothing to slow their development or deployment (Union of Concerned Scientists 2012). As yet, any agreement remains theoretical; the Conference on Disarmament, the primary international body for arms control, continues discussion on the draft treaty.

Perhaps more substantively, in 2010 the European Union publicly presented a draft Code of Conduct for Outer Space Activities, with signatories collectively responsible for preventing harmful interference or intentional damage to satellites. The draft represents a hypothetical agreement that space assets should no longer be a legitimate target of aggression. In January 2012, the United States announced that in lieu of signing the EU code, it would work with the European Union to develop an International Code of Conduct for Outer Space Activities (Space Code 2014). But would this Code of Conduct, if it came into force as a full-fledged treaty, solve the issues of dual use?

3.4 Dual-Use Conclusions

Unfortunately, no. Solving the issues of dual-use space technology by a voluntary code of conduct is unlikely to be successful. The advent of widespread dual-use technology in outer space is a terrifying, but potentially unavoidable, prospect. Once it becomes commonplace, space dual-use technology may incentivize first strikes on almost all space activities, a presumably unhappy outcome. In the meantime, we may be underestimating the threat. No nation has yet targeted an adversary's space assets; testing has so far always been on a nation-state's own space technology.

But without (yet) a first generation of sufferers, policymakers are likely to be overconfident that deployment of space weaponry, especially weaponry designed to be disguised as having a civilian intent, can avoid escalation to an unintended war. US Senator Ted Cruz (2017) has recently argued for space-based interceptors; such assets would be strategic targets for a first strike. And the strike need not be kinetic: disabling space capabilities will be likely to involve cyberattacks, particularly by nations (and non-state actors) without suitable kinetic weaponry. Unintended but foreseeable consequences are of notable concern. For example, China might view a strike on a space-based weapon platform as equally legitimate with a strike on similar ground-based weapons. The United States, however, might view the destruction of such a satellite as a prelude to nuclear strike. The result could be an unintended nuclear conflagration.

So where do we draw the lines between peaceful civilian use, merely potential dual-use, actual conflict short of war, acts of terrorism, and full-out war? We need a **red-line analysis**: how do we define and demarcate issues in space security along the spectrum from peaceful, clearly civilian interactions to armed conflict? Under what circumstances would satellite surveillance, or even weaponizing space, be (il)legitimate on the usual justifying ground of self-defense? Generally, what would conceivably legitimate a preventive or pre-emptive space war? A key point: whatever the correct answers to just war questions are on Earth, would they be the same in space? Answering that may shed light on when a space-based attack would constitute *casus belli*. One possible result is a requirement for the rigorous inspection and regulation of all launches and orbits, lest they become space weapons.

Analogies from other emerging technologies may also be instructive. For example, when do cyberattacks rise to the level of armed conflict or use of force (Allhoff and Jenkins 2014; Abney 2017b)? Or, what are the limits on "hacking back" as a defense against cyberattack (Lin 2016)? A major focus of the debate over "killer robots" on Earth has been the requirement for "meaningful human control," as groups like the Campaign to Stop Killer Robots (2018) have insisted on the prohibition of autonomous lethal robots. But no similar concerns have arisen over cyberconflict or conflict in space. Why? If "meaningful human control" is an ethical *sine qua non* in the killer-robot debate, why should it not also be a serious concern about cyberattacks, or the ethics of space war? After all, a lethal autonomous robot can kill in space just as easily as a human—in fact, more easily, given it needs no human life support systems, and will not have its performance degraded by weightlessness, radiation exposure, vertigo, nausea, muscle and bone atrophy, etc.

So, should we outlaw all possible autonomous space-based weapons, such as the "Star Wars" missile defense project? If defensive, but not offensive, autonomous space weapons are allowed, can we meaningfully distinguish between purely defensive and offensive weapons in space (e.g., Johnson-Freese 2007)? And for that matter, are there actually any purely defensive weapons in space?

4. Outer Space Bioethics and Technology Ethics

Scenario 2: In 2029, suppose you are part of the four-person crew on the first human spaceflight headed to Mars—on a SpaceX spaceship, launched on schedule in late 2028. Previous ships were already sent to build a basic habitat, and your ship is now five months away from landing. But something has gone terribly wrong. Micrometeorites have pierced the hull and caused a slow oxygen leak and radiation shielding failure; unless patched within two days, calculations show all four astronauts will die before landing. Patching the ship requires a spacewalk, but there is a solar storm raging that would give a lethal dose of radiation to a spacewalking astronaut. The crew is also already weakened by extended weightlessness and it is unclear if the two worst affected could even complete the repair. Choosing any crew member to die sacrifices essential mission-critical skills, making it a difficult dilemma to decide who gets to stay on the proverbial lifeboat. Suppose one of the crew is the designated mission commander if conflict breaks out with rival missions en route; must they be saved? Who should get to decide, and on what basis?

4.1 Human Health, Welfare, and Dual-Use Risk

The "lifeboat ethics" dilemma and related issues in space bioethics and technology has begun to be discussed (Lin and Abney 2014; Abney and Lin 2015; Abney 2017a). Such discussions build on previous approaches to space ethics based on religion and culture (Randolph et al. 1997; Peters 2013; Peters 2017) as well as broader ethical concerns (Arnould 2010; Persson 2012; Schwartz and Milligan 2016). Arguably, the previous dual-use discussion may also require a focus on space bioethics, insofar as manned missions and sample returns are planned, with related issues of planetary protection, especially backwards contamination.

NASA has studied the ethics and risks of long-duration human spaceflight (IOM 2014). But much has yet to be decided, starting with whether NASA rules should apply to private space enterprises. For example, in scenario 2, should each astronaut or other user of space be allowed to assess risks for themselves; or do we need some objective or third-party standard? What about involuntary or nonvoluntary risk versus voluntary risk? There already exists a track record of attempting to answer such questions about risk for other convergent technologies (Lin, Bekey, and Abney 2008; Abney 2012; Abney, Lin, and Mehlman 2013; Abney and Lin 2015; Abney 2017a; Abney 2017b; Abney and Ciupa 2018); do the different circumstances of space technology demand different answers?

What, for example, if a female astronaut is pregnant; should abortions in space be allowed, or even required, especially if they threaten the mission? Who would get to determine "acceptable risk" for a fetus, or more generally, for future generations? Further, how would the answers about proper risk assessment change if the ship is launched by NASA, instead of a private concern? What if it instead is launched by the Pentagon's new "Space Force," with the intention to defend the mission with kinetic weapons if anyone objects to an American colony on Mars? Generally, who will get to determine "acceptable risk" in space, and what method will they use for doing so? The answer will affect every conceivable use of space for the indefinite future.

Any ethical analysis or policy decision also assumes a concept of the moral community: who and what must we take into account in our decision-making, and who and what can safely be left aside as irrelevant? *What counts, and how do we count it?* So, what constitutes the moral community for space bioethics? Should we use a person-affecting (in which we only worry about current, actual persons) or a person-neutral ethics (Parfit 1984; Lin 2013)? And even for current persons, do we count them all equally? What would equality mean? For example, is it reasonable to make special accommodations for disabled people? Could those with traditional disabilities, for example blindness, even have an advantage in space (Wells-Jensen 2018)? Should there be special accommodations in crew selection or in other considerations for members of traditionally disadvantaged groups, for example, the LGBTQ community or pregnant women?

4.2 Space Bioethics and Dual-Use Technology

Space environments will pose serious dual-use bioethics concerns. In the second scenario, it takes little imagination to see pressure to militarize formerly civilian spaceflight and human experimentation; simply assume the spacecraft can be commandeered as a military vessel, and its commanding officer is (like most NASA pilots) from the military. Even for civilians, would the same regulations and protocols as on Earth (e.g., the Common Rule 1991) apply to human experimentation in space? Should that change for human experiments by the military, or private military contractors (Lin, Mehlman, and Abney 2013)? Whatever the decision-making body, quandaries abound when there is a grave risk or great uncertainty about risks, for example, prolonged exposure to radiation on a long mission or the long-term health of children born off-Earth.

Given that humans are all effectively disabled in space environments (Shew 2017), certain kinds of human enhancements (like radiation resistance, ability to function well at lower levels of oxygen, etc.) are foreseeably desirable in order to better adapt for travel and work in space. Could they pose any dual-use concerns either in space or upon return to Earth? That is, could an astronaut weaponize their own body (Abney and Lin 2015)? Most enhancements are irreversible; but an enhancement in space, such as reduced bone loss or better circulation in zero gravity, could well constitute a disability upon return to Earth. Should such enhancements be allowed?

Terrestrial life may not pose the only dual-use bioethics concern. Planetary protection is a long-standing worry—including both forward and back contamination (Meltzer 2010). For humans on Earth, the primary concern is back contamination—alien life (or alien technology) being returned to Earth and causing disease, even death. But how will we even know if we have encountered alien life? Can we distinguish it from heretofore unknown natural objects, or even possible alien artifacts? Can we be sure what is not (or no longer) alive? We need to assess the novel bioethical risks involved in the search for life, especially as regards sample retrieval, for example as planned in the Mars 2020 mission (Race et al. 2012; JPL 2018).

Even assuming we have safely captured alien life, more risks could ensue. It seems inevitable we would want to study the alien biota, not merely leave it alone in a completely isolated ecosystem. Who and what should determine risk and safety protocols for handling, retrieving, and experimenting upon alien life or artifacts (Rummel et al. 2002)? Under what circumstances (if any) should such discoveries be classified or kept secret, or deemed too great a risk for sample return? Should off-Earth facilities (e.g., on the ISS) and quarantine protocols be used for research on any newly discovered extraterrestrial life until safety has been established (Abney 2017a), following a precautionary principle? Under what circumstances (if any) should we deny living astronauts the opportunity to return to Earth (Abney and Lin 2015)? It is conceivable that we may need to invoke, or update, the Biological and Chemical Weapons Conventions for policies regarding newly discovered extraterrestrial life.

4.3 Colonization and Existential Risk

Elsewhere (Lin and Abney 2014; Abney and Lin 2015; Abney 2017a) I have discussed in more detail the bioethics of the discovery and exploration of space by our technology versus going ourselves in person, arguing that the risks make human spaceflight unethical in most circumstances. Here, I turn here to two possible exceptions that could potentially satisfy space-based bioethics concerns, ones that have begun to garner substantial attention: colonization and existential risk.

First, is there a moral imperative to colonize space (Smith 2016)? Regardless, many consider colonization inevitable; Gregory Cooper (2016) bets that humans will colonize the entire galaxy within a million years, based on what he considers reasonable projections about technological development in terraforming and spaceflight. Assuming humanity eventually becomes a multi-planet species, who should decide how Mars or other bodies in space are adapted for human habitation—private enterprise, governments, or some third option? Some degree of changing the environment is inevitable for human habitation and survival; so, what degree of terraforming is permissible (York 2002)? For one example of terraforming's dual-use concerns, should Elon Musk's proposal (2015) to use nuclear weapons to hasten climate change on Mars be permitted? If not, who should, or could, stop him?

Let's begin an assessment by addressing some practical objections to colonization. First, there are few reasons to think short to medium term attempts at human colonization of Mars (or the Moon, or anywhere else off-world) would be successful. Microgravity, micrometeorite impacts, radiation, or even the stresses of isolation and confinement may kill astronauts en route to Mars–or cause them to kill each other–during their long journey (Abney and Lin 2015). And even if the would-be settlers could get there in one piece, it seems likely they would succumb soon thereafter (Do et al. 2014). Given "ought implies can," if success in colonization is practically impossible, then it makes no sense to say we ought to do it. This problem is exacerbated in realizing that the success of colonies would depend not just on individual survival, but reproduction; and the odds of successfully having babies on Mars are even longer than simply getting there alive. Even the attempt to reproduce in such dangerous circumstances would quite possibly violate numerous bioethical precepts (Lin and Abney 2014), as such experiments on fetuses would never make it past an IRB on Earth.

But Musk's argument (2017) for the importance of colonization does not deny these practical concerns; instead, he ties together ethical issues concerning colonization and the topic of existential risk, so it is worth examining their mutual relationship to technology more closely. First, let's define existential risk (Bostrom 2002, 15): "the chance that an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential." Most anyone who thinks about the topic agrees that existential risk is a Very Bad Thing; but *exactly* how bad? Specifically, does existential risk always trump all other ethical and risk considerations (Bostrom 2013; Abney 2017a)? If not, how do we do the relative risk assessment?

Our own actions in space could raise existential risk; for example, consider intentional messaging to extraterrestrial intelligences (METI). Enthusiasts for METI (e.g., Vakoch 2017; Kitchen 2018) plan on direct, intentional messaging to specific astronomical objects in the hope of alerting possible intelligent extraterrestrials to our presence. But some warn that the galaxy may be full of menacing silent hunters, termed the "Dark

Forest" theory (Liu 2008), and others similarly argue for due diligence before shouting into the cosmos (Brin 2014). Should we require informed consent for what effectively is an experiment affecting all humans (Smith 2019)? If METI is permitted, should distance matter? (Given the laws of physics, messaging a civilization 7,000 light-years away presumably poses a far more distant threat in time as well as space than one 4.3 light-years away). Even if METI is permitted, should there be restrictions on its message content? And if alien life is discovered, does existential risk only apply to humanity? Could protection of alien life ever be worth sacrificing our own?

Musk (2017) and others explicitly endorse colonization on the basis that it mitigates existential risk, giving us a "backup planet" in case life meets a catastrophe on Earth (cf. Bostrom 2016). Does that argument justify such efforts, or does it merely miss the point? After all, if we will simply ruin every place we settle, isn't colonizing Mars merely postponing the inevitable (Lin 2006)? Critics such as Lori Marino (2018) point out that Musk, Bezos, and other space-mad billionaires have done very little to protect Earth's ecosystems, stop (or even mitigate the effects of) climate change, or much of anything else to make Earth more livable in the long term. If we don't change our ways, then we will export our problems to Mars as well, and so colonization will not even really address existential risk!

Is it then only morally acceptable to colonize Mars once we have made life on Earth sustainable indefinitely (Marino 2018), so that we know how to do the same on Mars? If so, doesn't that undermine the entire point of near or medium-term colonization as a mitigation of existential risk? Not even to mention that, with near-term technology, it almost certainly won't work (Do et al. 2014). There are grave risks to sending humans on a quixotic quest to populate the cosmos, ones that diminish if we give up on sending humans, and instead persist in sending only robots, to space. The risks seem to outweigh most justifications for colonization. But mitigating existential risk may be the one justification that would change the risk-benefit assessment for even small odds of successful colonization. The details of that argument are next.

4.4 Why Is Mitigating Existential Risk so Important—and Colonization the Solution?

To explain, we first need to explore some basic ethics. There are three common basic approaches to ethics: consequentialism, deontology, and virtue ethics (Abney 2012). To oversimplify, a consequentialist maintains that right action is whatever would produce the best consequences; the end justifies the means. Of course, there's an epistemic issue: we don't know which interventions will work for certain—we just have estimates. So consequentialists routinely appeal to the concept of expected utility: multiply the probability times the magnitude of the good consequences, then subtract the probability times the magnitude of the bad consequences. The result is the expected utility; one should maximize that (Abney 2018).

What is utility's relevance to space colonization and existential risk? Well, suppose we discover a killer asteroid too late to stop its impact on Earth, or fall prey to any of a number of other potential calamities that could imperil civilization. If a permanent, sustainable colony off-world already existed (as remote as that prospect currently appears), then humanity could survive such a cataclysm on Earth, and potentially spread throughout the cosmos. Accordingly, utilitarian existential-risk theorists, such as effective altruists thinking about the long-term future (Whittlestone 2017), hold that colonization efforts might be worthwhile, even if the odds of success are minuscule. We will always maximize expected utility by minimizing existential risk, which effectively adds some uncountably large number of future humans to our equation. To see why, do the math: (close to) infinity future humans times any small percentage always outweighs any finite number times even a large percentage. So, if colonization thereby minimizes existential risk, then a standard consequentialist account would imply that we should spend more money on it than anything else–even if the odds of success are low.

What about deontology, then? Deontologists routinely distinguish between *prima facie* (sometimes termed *pro tanto*) duties, which hold unless they are overridden by some other, competing duty; versus absolute duties, which hold no matter what. It can sometimes be ethical to violate a *prima facie* duty, if upholding it would violate some other, equally or even more important duty. But it is always unethical to violate an absolute duty; it takes precedence over every other obligation one could have.

So understanding an absolute duty is crucial to ethics—if any exist. Various moral theories claim they do, but differ as to what they are. The most plausible way of justifying that a duty is absolute is to argue that it is required for morality itself to exist (Abney 2018). That is, any duty that conflicted with such an absolute duty could not be ethically required, because it would do away with ethical requirements! What kind of duty could itself be morally required for morality itself to exist? Well, both I (e.g. Abney 2018; Abney 2019) and Brian Green (2018), following the work of Hans Jonas, have argued that ensuring humanity's continued existence is such an absolute duty.

If so, we can formulate a plausible absolute duty: the *Extinction Principle* (Abney 2019): "one always has a moral obligation never to allow the extinction of all creatures capable of moral obligation." It then is an absolute duty to keep things capable of obeying absolute duties in existence. Accordingly, mitigating existential risk is an absolute duty, which wins any conflict it has with any other duty. If space colonization is the activity that most minimizes existential risk, then it is our highest duty.

Virtue ethics may yield a different emphasis than deontological or utilitarian approaches; it's plausible that a virtue ethicist might insist that an obsession with decreasing existential risk to the detriment of other aspects of human flourishing betrays a flawed, even vicious character. But for the deontological *Extinction Principle* or a standard version of expected utility, decreasing existential risk trumps all other considerations. And, plausibly, one of the best ways of reducing existential risk is to make humans into a multiplanet species (Musk 2017). So sending humans, not just robots, into space may be crucial for decreasing existential risk.

4.5 What's the Hurry? The Interstellar Doomsday Argument

Even though deontologists and utilitarians may agree that reducing existential risk is morally crucial, and space colonization perhaps the best means to do so, they still might not think it a pressing priority now. The reason, as alluded to, is the low odds of success in the short term. Perhaps an effective altruist or deontologist fond of the Extinction Principle might instead argue that we should amply fund basic and applied research on what it takes to create a self-sustaining colony off-world, and until that research has matured, attend to other, more terrestrial concerns.

The underlying presumption of such an approach is that it is overwhelming likely that existential catastrophe will not happen soon; we should take plenty of time to get colonization right before we actually attempt it, so we should wait until the odds of success are higher; perhaps until they are over 50 percent.

But from an existential risk perspective, it is dangerous to wait. If we are to save humanity by becoming a multiplanet species, we may need to start very, very soon. The Interstellar Doomsday Argument gives one reason why, directly connected to our level of technology. Here is a brief version (for more details, see (Abney 2017a, Abney 2019)):

> First, the "Self-Sampling Assumption" (SSA): "One should reason as if one were a random sample from the set of all observers in one's reference class" (Bostrom and Cirkovic 2011).
>
> Next, a data point: our first robotic envoy to the stars, Voyager 1, entered interstellar space in August 2012 (Cook and Brown 2013).
>
> Next, apply the SSA: assume you are a random observer within a species that has achieved interstellar travel. As of the publication of this text, that was about eight years ago.
>
> Next, Gott's (1993) "delta t argument": expect a 95 percent probability that any randomly observed phenomenon will continue to exist for between 1/39 and 39 times its present age (termed "L"), given a 5 percent possibility your random observation comes in either the first or last 2.5 percent of its lifetime.
>
> Conclusion: there is a 95 percent chance that our future as a species with interstellar probes will only last between L/39 (75 more days) and 39L (312 more years).

Further, there is a 75 percent chance that our future as a species with interstellar probes will last less than 3L—24 more years.

4.6 Doom Soon?

The full reason this pessimism is justified constitutes the remainder of the Interstellar Doomsday Argument.

Consider Fermi's paradox: if aliens exist, why isn't their existence obvious? That is, "where is everybody"? David Brin (1983) reformulated this as the "Great Silence": if aliens exist, why don't we see clear evidence of their presence in the cosmos? Why are they silent?

Next, the Drake equation, which calculates the number of detectable alien civilizations currently in the galaxy, N, as $N = R^{*} \bullet fp \bullet ne \bullet fl \bullet fi \bullet fc \bullet L$ (SETI 1961).

Next, Robin Hanson (1998) postulates the "Great Filter." It explains the Great Silence by one (or more) of the as yet unknown variables in the Drake equation having a near-zero value.

There is increasing consensus in the astronomical community that none of the first three variables are close to zero (Seager 2016). So, to explain the Great Silence, one or more of the last four, biological, factors in the Drake equation must approach zero. It may be one of the first three biological variables have a cosmic value near zero, making intelligent, communicating life here on Earth an evolutionary miracle; then the Great Filter is in our past. But if many past civilizations have arisen and developed detectable interstellar technology, then the Great Silence seemingly implies that L is close to zero: the Great Filter is in our future. Whenever previous alien civilizations ascended to our current level of technology, they became undetectable very, very quickly. The most plausible way to render our civilization undetectable very soon is, of course, human extinction (Abney 2017a). (It seems unlikely everyone would agree to give up *all* detectable technology and yet have us survive long as a species.)

And that is particularly true of one technology, which is the most practical, relatively inexpensive, and longest-lasting way to be detected across the galaxy for millions, even billions, of years—our robotic spacecraft. If other past civilizations in the Milky Way's 13 billion year history did send robotic probes into interstellar space, just as we now have, it seems (overwhelmingly) likely that some of those probes would be here by now. Take von Neumann probes, that is, probes capable of self-reproduction. (Whether such probes would count as life is an interesting question.) Such a probe could, upon arrival at its target destination, use 3D printing and materials found in situ to produce copies of itself, and then send those copies on to other stars (Freitas 1980).

Using extremely conservative assumptions, Stephen Webb (2002) estimates saturation of every solar system in the galaxy by at least one von Neumann probe would take at most 4 million years, less than 1/3000th of the age of the Milky Way. Even if one assumes *all* other civilizations eschew von Neumann probes and stick to our slow, non-reproducing approach to interstellar robotic technology, 13 billion years is still plenty of time to fill the Milky Way with robots like our Voyagers, launched from alien homeworlds. And this doesn't even consider programs like Yuri Milner's Breakthrough Starshot, based on concepts that would greatly accelerate interstellar robotic exploration (Lubin 2016). Simply put, given that our Solar System is apparently bereft of robotic probes from ancient alien civilizations, it seems incredible to believe that they're on the way, and just need more time to get here.

Now, one might argue that we know more than the mere fact that a robot hit interstellar space eight years ago—for Voyager 1 has since been joined by four more spacecraft on escape trajectories from the solar system, along with three additional rocket motors on interstellar trajectories, all of which would constitute convincing evidence to aliens of our civilization (Johnston 2015); and more are planned. But that merely reinforces the point; we have no reason to believe this moment in time is privileged with respect to our robotic interstellar probes. We reached the threshold eight years ago, and there's no sign we're going to stop. But, clearly, it has stopped (or never started), everywhere else; that is the message of the Great Silence.

And let's be clear about the technology: sending robots to the stars is vastly easier than having humans colonize Mars, much less any other planet or moon or artificial construction, like an O'Neill cylinder (Abney 2019). We can barely envision a sustainable off-world colony for humans—but we already have interstellar robotic emissaries. If we become incapable of sending such probes very, very soon, then presumably we humans will be unable to escape the Earth. And if we cannot escape the Earth, then sooner or later we will go extinct. (My bet is sooner.) So, this argument should reinforce a sense of urgency: if humans are to escape becoming just another species to go extinct on the Earth, whether through external calamity or self-inflicted wounds, we had better get our colonists, and not just our robots, off-planet. Unless, of course, it is precisely that attempt at setting up off-world colonies, as opposed to sustainable tending of their own terrestrial gardens, that doomed all the other civilizations?

5. Conclusion

I believe the colonization imperative to avoid existential risk is the only serious moral argument for encouraging a human presence in space; not the vanity missions of insisting a human set foot on Mars, or Ganymede, or Ceres, or . . . well, it's a short list where humans could even conceivably set a space suited boot off-Earth. Robots can do it better, faster, cheaper, and their edge will only grow as robotic technology advances. The only good reason to send humans rather than mere tech to space is for colonization, and in the short to medium term, such efforts are almost guaranteed to fail. Dual-use issues raise further concerns for human activity in space; future treaties will require monitoring that will be considerably complicated if we allow human astronauts to interact with ostensibly civilian spacecraft. Hence, for the foreseeable future, I conclude that sending humans on space missions is simply immoral, until and unless our technology has prepared the way for our colonists to succeed. As space exploration and exploitation become more common, we must consider now how best to guide humanity and its technology responsibly and reflectively into the cosmos. Otherwise, as is the human wont, we may have to find out the answers the hard way, by making a first generation suffer. But the first generation problem in space could conceivably become a last generation problem as well; for issues of existential risk, discovering the answers too late may mean never discovering them at all, as the human experiment comes to an end.

References

Abney, Keith. 2004. "Sustainability, Morality and Future Rights." *Moebius* 2, no. 2. http://digitalcommons.calpoly.edu/moebius/vol2/iss2/7/

Abney, Keith. 2012. "Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by P. Lin, K. Abney, and G. Bekey. Cambridge. MA: MIT Press.

Abney, Keith. 2017a. "Robots and Space Ethics." In *Robot Ethics 2.0*, edited by P. Lin, R. Jenkins, and K. Abney, 354–368. New York: Oxford University Press.

Abney, Keith. 2017b. "On the Ethics of Cyberwar." *Communications of the ACM*, July 26, 2017. https://cacm.acm.org/blogs/blog-cacm/219696-on-the-ethics-of-cyberwar/fulltext

Abney, Keith. 2018. "On Aliens and Robots: Moral Status, Epistemological and (Meta-)ethical Considerations." Presentation to *Social and Conceptual Issues in Astrobiology Conference* (SoCIA 2018), University of Nevada, Reno, April 13, 2018.

Abney, Keith. 2019. "Ethics of Colonization: Arguments from Existential Risk." *Futures*, Volume 110: 60–63. doi:10.1016/j.futures.2019.02.014.

Abney, Keith, and Martin Ciupa. 2018. "AI Conceptual Risk Analysis Matrix (CRAM)." Presentation at *International Conference on Robot Ethics and Standards* (ICRES 2018), Troy, New York, August 20–21, 2018.

Abney, Keith, and Patrick Lin. 2015. "Enhancing Astronauts: The Ethical, Legal, and Social Implications." In *Commercial Space Exploration: Ethics, Policy, and Governance*, edited by Jai Gaillott. New York: Ashgate.

Allhoff, Fritz, and Ryan Jenkins. "When Is a Real-World Response to a Cyberattack Justifiable?" *Slate*, June 11, 2014. http://www.slate.com/articles/technology/future_tense/2014/06/cyberwar_ethics_when_is_a_real_world_response_to_a_cyberattack_justifiable.html

Arnould, Jacques. 2009. "Astrobiology, Sustainability, and Ethical Perspectives." *Sustainability* 1, no. 4: 1323–1330.

Arnould, Jacques. 2010. "Space Ethics." In *Space Exploration and Humanity: A Historical Encyclopedia*, edited by S.B. Johnson, 1071–1072. Santa Barbara: ABC-CLIO.

Axe, David. 2014. "The Pentagon's Plan to Put Robot Marines in Space." *The Week*, July 9, 2014. https://theweek.com/articles/445664/pentagons-plan-robot-marines-space

Axe, David. 2016. "Russia Is Building a Nuclear Space Bomber." *The Daily Beast*, July 14, 2016. http://www.thedailybeast.com/articles/2016/07/14/russia-is-building-a-nuclear-space-bomber.html

Bostrom, Nick. 2002. "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards." *Journal of Evolution and Technology* 9 (1). http://www.nickbostrom.com/existential/risks.html

Bostrom, Nick. 2013. "Existential Risk Prevention as Global Priority." *Global Policy*, 4, no. 1 (2013): 15–31.

Bostrom, Nick. 2016. "The Existential Risk FAQ." http://www.existential-risk.org/faq.html

Bostrom, Nick, and Milan M. Cirkovic. 2011. *Global Catastrophic Risks*. New York, NY: Oxford University Press.

Boyle, Rebecca. 2011. "Who Owns the Moon's Water? Future Moon Mining Missions May Face Le-gal Disputes." *Popular Science*, January 19, 2011. https://www.popsci.com/science/article/2011-01/moon-miners-would-need-good-lawyers-shore-extracted-resources

Brin, G. David. 1983. “The ‘Great Silence’: The Controversy Concerning Extraterrestrial Intelligent Life.” *Quarterly Journal of the Royal Astronomical Society*, vol. 24, no. 3, 283—309.

Brin, G. David. 2014. “The Search for Extraterrestrial Intelligence (SETI) and Whether to Send ‘Messages’ (METI): A Case for Conversation, Patience and Due Diligence.” *Journal of the British Interplanetary Society*, 67, 8–16.

Cain, Fraser. 2017. “Can We Launch Nuclear Waste into the Sun?” *Universe Today*, Feb. 27, 2017. https://www.universetoday.com/133317/can-we-launch-nuclear-waste-into-the-sun/

Campaign to Stop Killer Robots. 2018. https://www.stopkillerrobots.org/learn/

Chao, Tom. 2014. “3D Printing: 10 Ways It Could Transform Space Travel.” Space.com. http://www.space.com/25706-3d-printing-transforming-space-travel.html

Common Rule. 1991. “Federal Policy for the Protection of Human Subjects (‘Common Rule’).” https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html

Cook, Jia-Rui C., and Dwayne Brown. 2013. “NASA Spacecraft Embarks on Historic Journey Into Interstellar Space.” NASA, September 12. https://www.nasa.gov/mission_pages/voyager/voyager20130912.html.

Cooper, Gregory. 2016. “Within 1 Million Years, Humanity or Its Descendants Will Have Colonised the Galaxy.” *Long Bets: The Arena for Accountable Predictions*. http://longbets.org/721/

Cornish, Chloe. 2017. “Interplanetary Players: A Who’s Who of Space Mining.” *Financial Times*, October 18, 2017. https://www.ft.com/content/fb420788-72d1-11e7-93ff-99f383b09ff9

Cruz, Ted. 2017. “How to Degrade the Growing Power of North Korea.” *Washington Post*, August 1, 2017. https://www.washingtonpost.com/opinions/how-to-degrade-the-growing-power-of-north-korea/2017/08/01/64843812-7609-11e7-8839-ec48ec4cae25_story.html?utm_term=.997191916ad1

Do, Sydney, Koki Ho, Samuel Schreiner, Andrew Owens, and Olivier de Weck. 2014. “An Independent Assessment of the Technical Feasibility of the Mars One Mission Plan—Updated Analysis.” *Acta Astronautica* 120 (March–April): 192–228.

Dorminey, Bruce. 2017. “Trump Should Make Space-Based Solar Power A National Priority.” *Forbes*, March 18, 2017. https://www.forbes.com/sites/brucedorminey/2017/03/18/trump-should-make-space-based-solar-power-a-national-priority/#1db5544a3e69

Freitas, Robert A., Jr. 1980. “A Self-Reproducing Interstellar Probe.” *Journal of the British Interplanetary Society* (July 1980) 33: 251–264.

Gingrich, Newt. 2012. Campaign speech in Cocoa, FL, on January 25, 2012. https://abcnews.go.com/Technology/newt-gingrich-promises-moon-base-flights-mars-reality/story?id=15449425

Gott III, J. Richard. 1993. “Implications of the Copernican Principle for Our Future Prospects.” *Nature* 363, no. 6427: 315–319.

Gravity (movie). 2013. Directed by Alfonso Cuarón. https://www.warnerbros.com/gravity

Green, Brian. 2018. “Self-Preservation Should Be Humankind’s #1 Ethical Priority and Therefore Rapid Space Settlement Is Necessary.” *Futures* 110, 35–37.

Hanson, Robin. 1998. “The Great Filter—Are We Almost Past It?” George Mason University, September 15. https://mason.gmu.edu/~rhanson/greatfilter.html

Hardin, Garrett. 1968. “The Tragedy of the Commons.” *Science* 162, no. 3859 (December 13): 1243–1248.

IOM, Institute of Medicine Committee on Ethics Principles and Guidelines for Health Standards for Long Duration and Exploration Spaceflights. 2014. “Health Standards for Long Duration and Exploration Spaceflight: Ethics Principles, Responsibilities, and

Moon Treaty. 1979. "Agreement Governing the Activities of States on the Moon and Other Celestial Bodies." http://www.unoosa.org/oosa/en/ourwork/spacelaw/treaties/intromoon-agreement.html

Musk, Elon. 2015. Interview on *The Late Show* with Stephen Colbert, September 10, 2015. https://www.youtube.com/watch?v=gV6hP9wpMW8

Musk, Elon. 2017. "Making Life Multiplanetary." Presentation at 68th annual International Aeronautical Congress, September 28, 2017, Adelaide, Australia. http://www.spacex.com/sites/spacex/files/making_life_multiplanetary_transcript_2017.pdf

Nielsen, Nick. 2014. "The Infrastructure Problem." Centauri Dreams. http://www.centauri-dreams.org/?p=30525

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Obama, Barack. 2015. "U.S. Commercial Space Launch Competitiveness Act (H.R. 2262)." https://www.planetaryresources.com/2015/11/president-obama-signs-bill-recognizing-asteroid-resource-property-rights-into-law/

Outer Space Treaty. 1967. U.S. Department of State. https://www.state.gov/t/isn/5181.htm

Palaszewski, Bryan A. 2014. "Atmospheric Mining in the Outer Solar System: (Aerial Vehicle Reconnaissance and Exploration Options)." NASA Technical Reports Server. http://ntrs.nasa.gov/search.jsp?R=20140017392

Parfit, Derek. 1984. *Reasons and Persons*. New York: Oxford University Press.

PAROS draft treaty. 2017. http://www.nti.org/learn/treaties-and-regimes/proposed-prevention-arms-race-space-paros-treaty/

Persson, Erik. 2008. *What Is Wrong with Extinction?* Lund: Lund University.

Persson, Erik. 2012. "The Moral Status of Extraterrestrial Life." *Astrobiology* 12(10): 976–984.

Peters, Ted. 2013. "Astroethics: Engaging Extraterrestrial Intelligent Life-Forms." In *Encountering Life in the Universe*, edited by C. Impey, A. Spitz, and W. Stoeger, 200–221. Tucson: University of Arizona Press.

Peters, Ted. 2017. "Astrobiology, Astrotheology, and Astroethics." International Society for Science and Religion. https://www.issr.org.uk/blog/astrobiology-astrotheology-astroethics-ted-peters/

Plait, Phil. 2007. "Why Explore Space?" *Bad Astronomy*, November 28, 2007. http://blogs.discovermagazine.com/badastronomy/2007/11/28/why-explore-space/.

Plait, Phil. 2012. "The Newtonian Mechanics of Building a Permanent Moon Base." *Bad Astronomy*, January 27, 2012. http://blogs.discovermagazine.com/crux/2012/01/27/the-newtonian-mechanics-of-building-a-permanent-moon-base/

Planetary Resources. 2018. "Why Asteroids?" https://www.planetaryresources.com/why-asteroids/

Posey, Bill. 2014. "Space: The Ultimate High Ground." *Space News*, February 24, 2014. https://spacenews.com/39613space-the-ultimate-high-ground/

Race, Margaret, Kathryn Denning, Constance M. Bertka, Steven J. Dick, Albert A. Harrison, Impey, Christopher Harrison, Rocco Mancinelli, and workshop participants. 2012. "Astrobiology and Society: Building an Interdisciplinary Research Community." *Astrobiology* 12, no. 10: 958–965.

Randolph, Richard, Margaret Race, and Christopher P. McKay. 1997. "Reconsidering the Theological and Ethical Implications of Extraterrestrial Life." *The Center for Theology and the Natural Sciences Bulletin* 17: 1–8.

Rodriguez, Julie. 2013. "Japan Plans to Harvest Solar Power from Space by 2030." *Inhabit*. http://inhabitat.com/japan-plans-to-harvest-solar-power-from-space-by-2030/

Rummel, John D., Margaret S. Race, Donald L. DeVincenzi, P. Jackson Schad, Pericles D. Stabekis, Michel Viso, and Sara E. Acevedo, eds. 2002. "A Draft Test Protocol for Detecting Possible Biohazards in Martian Samples Returned to Earth." NASA/CP-2002-211842. Washington, D.C.: US Government Printing Office.

Schneider, Jean. 2013. "Philosophical Issues in the Search for Extraterrestrial Life and Intelligence." *International Journal of Astrobiology* 12: 259–262.

Schwartz, James. 2016. "Near-Earth Water Sources: Ethics and Fairness." *Advances in Space Research* 58: 402–407.

Schwartz, James, and Tony Milligan, eds. 2016. *The Ethics of Space Exploration*. Switzerland: Springer.

Seager, Sara. 2016. "Research." MIT website. http://seagerexoplanets.mit.edu/research.html

SETI Institute. 1961. "The Drake Equation." SETI Institute website. http://www.seti.org/drakeequation.

Shew, Ashley. 2017. "Technoableism, Cyborg Bodies, and Mars." https://techanddisability.com/2017/11/11/technoableism-cyborg-bodies-and-mars/

Simberg, Rand. 2012. "Property Rights in Space." *New Atlantis*, Fall 2012. https://www.thenewatlantis.com/publications/property-rights-in-space

Singer, Peter. 1974. "All Animals Are Equal." *Philosophic Exchange* 5, no. 1: Article 6.

Singh, Timon. 2011. "Japanese Corporation Plans to Turn the Moon into a Massive Solar Power Plant." *Inhabit*. http://inhabitat.com/japanese-corporation-plans-to-turn-the-moon-into-a-massive-solar-power-plant/

Skibba, Ramin. 2018. "Mining in Space Could Lead to Conflicts on Earth." *Nautilus*, May 2, 2018. http://nautil.us/blog/-mining-in-space-could-lead-to-conflicts-on-earth

Smith, Kelly. 2016. "Cultural Evolution and the Colonial Imperative." In *Dissent, Revolution, and Liberty Beyond Earth*, edited by C. Cockell, 169–187. Switzerland: Springer.

Smith, Kelly. 2019. "Homo Reductio: Defusing a Radical Objection to Human Colonization of Other Worlds." *Futures* 110. 10.1016/j.futures.2019.02.005.

Snyder, Mike. 2014. "Welcoming the Era of In-Space Manufacturing." Space.com. http://www.space.com/27870-3d-printer-made-in-space-op-ed.html

Space Code. 2014. "The International Code of Conduct for Outer Space Activities" (last public draft 2014). http://www.eeas.europa.eu/non-proliferation-and-disarmament/pdf/space_code_conduct_draft_vers_31-march-2014_en.pdf).

Stilwell, Blake. 2018. "The Air Force's "Rods from God" Could Hit with the Force of a Nuclear Weapon—With No Fallout." Business Insider.com. https://www.businessinsider.com/air-force-rods-from-god-kinetic-weapon-hit-withnuclear-weapon-force-2017-9?IR=T

Union of Concerned Scientists. 2012. "A History of Anti-Satellite Programs." February 2012. http://www.ucsusa.org/nuclear-weapons/space-security/a-history-of-anti-satellite-programs)

Vakoch, Douglas A. 2017. "Good Call." *The New Scientist* 236(3154): 24–25.

Veruggio, Gianmarco, and Abney, Keith. 2012. "Roboethics: The Applied Ethics for a New Science." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by P. Lin, K. Abney, and G. Bekey. Cambridge: MIT Press.

Webb, Stephen. 2002. *If the Universe Is Teeming with Aliens … Where Is Everybody?* New York: Copernicus Books.

Wells-Jensen, Sheri. 2018. "Things You Didn't See Because You Were Looking: Blind Aliens, Science, and Interspecies Miscommunication." Presentation to Social and Conceptual Issues in Astrobiology Conference (SoCIA 2018), University of Nevada, Reno, April 13, 2018.

Whittington, Mark. 2015. "How Newt Gingrich's Moon Base Became 'Pretty Cool.'" *The Hill*. http://thehill.com/blogs/congress-blog/technology/257501-how-newt-gingrichs-moon-base-became-pretty-cool

Whittlestone, Jess. 2017. "The Long Term Future." *Effective Altruism*. https://www.effectivealtruism.org/articles/cause-profile-long-run-future/

York, Paul. 2002. "The Ethics of Terraforming." *Philosophy Now* 38, October/November. https://philosophynow.org/issues/38/The_Ethics_of_Terraforming

PART VII

TECHNOLOGY AND THE GOOD LIFE

CHAPTER 28

TECHNOLOGY, COGNITIVE ENHANCEMENT, AND VIRTUE ETHICS

BARBRO FRÖDING

1. Introduction

This chapter explores how cognitive enhancement, by means of technology, in combination with a commitment to virtue ethics could improve our capacity for responsible decision making. Such decision making includes epistemic as well as moral components. It involves the ability to think and act in a way that is conducive to the well-being of both the individual and the collective.

Section 2 explains what is meant by enhancement in this context and offers some relevant distinctions, such as that between pharmacological and non-pharmacological methods of cognitive enhancement. Section 3 lists some key cognitive shortcomings of humans which, plausibly, have a negative impact on our capacity for making well-informed, responsible decisions. The fourth section consists of a very brief introduction to some aspects of virtue ethics and sketches how properly instilled epistemic and moral virtues are conducive to good decision making. Section 5 provides four examples of non-pharmacological cognitive enhancement technologies: (i) computer training, (ii) neurofeedback or electroencephalogram (EEG) biofeedback, (iii) transcranial direct stimulation (tDCS,) and (iv) brain-computer interface (BCI) and looks at their (potential) effects on some core cognitive capacities. Section 6 discusses the alleged tensions between cognitive enhancement and the virtue ethical tradition of education and habituation as the primary means to instill good behavior. It is suggested that the two are, in many cases, complementary and indeed necessary for the good life. The chapter finishes with some general comments on ethical aspects of human cognitive enhancement which are present regardless of whether we use pharmacological or non-pharmacological means of enhancement.[1]

2. Enhancement

Human enhancement can be defined as "*biomedical interventions that are used to improve human form or functioning beyond what is necessary to restore or sustain health.*" (Juengst and Moseley 2016). While this and similar definitions fail to create a sharp distinction between treatment (including preventive treatment) and enhancement, it provides a broad understanding for the type of practices discussed in this chapter (for a further discussion see Daniels 2007; Bostrom and Roache 2008).

There are many forms of enhancement. Examples include physical enhancement, mood enhancement, lifespan extension, moral enhancement and, the subject of this chapter, enhancement of cognitive skills. Very broadly speaking one can split the methods for achieving cognitive enhancements into two groups; pharmacological and non-pharmacological. The latter group includes both conventional methods such as education, improved health (e.g., sleep), coffee, mental training (e.g., focused attention training), omega acids (Luchtman and Song 2013), and cardiovascular exercise, as well as the more unconventional methods; for example, transcranial magnetic stimulation (TMS), BCI; as developed by, for example, Facebook, Neuralink and Kernel), and genetic modifications (Dresler et al. 2013). For the last two decades most of the debate on cognitive enhancement has focused on pharmaceuticals and until recently enhancement by means of other technologies has received less interest. A contributing reason is that most of the technology is still in its early stages and much of the discussion inevitably becomes highly speculative. That said, early identification and discussion of the ethical challenges of emerging technologies before they are upon us is, of course, prudent.

For this chapter I have purposely selected examples of non-pharmacological technologies which, although far from fully developed, at least are in existence. The first example is various types of computer training. This might seem like a mundane and generally unexciting technology in this context. That is exactly the point: I want to anchor the discussion in technology which is available now. Additionally, consider some of the advantages of computer training (in comparison to other technologies): for example, wide availability, low price, low risk, and potentially high uptake.

In addition to staying with "current" technology, I have chosen to discuss cognitive effects which are still within the range of what would be considered normal for humans. This can be contrasted with so called "radical" enhancements or supra-enhancement: for example, eternal life, super-human intelligence, a new sense, or a novel mental capacity (Kahane and Savulescu 2015). It is sometimes assumed that "normal range" enhancements would not have a substantial effect. That would be a mistake. As pointed out by Guy Kahane and Julian Savulescu: "But even changes that operate within the currently normal range can be dramatic. It would be dramatic enough if an intervention gave most people an IQ of 140, or a lifespan of 110, even if both figures are well within the normal range. And even interventions that just increase or reduce the current diversity of dispositions and capacities might, in some context, be very important" (2015, 143).

As observed by Nick Bostrom and Anders Sandberg an important aspect of cognitive enhancements is that "they improve *core cognitive capacities* rather than merely particular narrowly defined skills or domain-specific knowledge" (Bostrom and Sandberg 2009, 312). So although less spectacular than supra-enhancement, an improvement with regards to, for example, working memory, focused and sustained attention, cognitive flexibility, and learning could have a positive impact on our capacity for the type of responsible, reflective, and rational decision making we need more of. Arguably, improved core cognitive skills could make a person more aware of her own pre-understanding, more able to compute fragmented, contradictory and complex information, as well as distinguish information from disinformation. This could have a favorable impact on, for example, our capacity for risk assessment. It must immediately be added, however, that improved cognitive capacities in no way guarantee neither willingness nor ability to act morally. It is easy to imagine an enhanced and vicious agent using their new cognitive skills to profit at other people's expense. This will be discussed in Section 6 of this chapter, where I argue that a virtue ethical framework could normatively anchor the cognitive skills and increase the likelihood of morally good behavior.

3. Cognitive Shortcomings and Increasing Demands

The rapid technological developments and the ever accelerating flows of information and disinformation in the twenty-first century make increasing cognitive and emotional demands upon us. In this complex information environment it becomes ever more challenging to make well-informed, responsible decisions. Unfortunately, scientific research has shown that humans might be less able to tackle these demands than previously thought. Studies in, for example, neurophysiology and neuropsychology have shown that stress and information overload have a very negative impact on our memory, our capacity for risk-assessment, as well as epistemic deference; they also undermine self-control (Arnsten 2009; Qin et al. 2009; Selart and Johansen 2011). Stress also increases tendencies towards experiential avoidance and has a negative effect on emotional regulation which, in turn, prolongs the stress (Wegner et al. 1987; Hayes et al. 1996; Golkar et al. 2014). Other studies in moral psychology and behavioral economics have shown the negative impact of systemic bias. Systemic bias is used here as an umbrella term that includes, for example, status quo bias and confirmation bias. It is the inherent tendency of a (mental) process primed by biological and social/environmental factors, which influences behavior and decision making. Notably, we tend to be unaware of how much it impedes on our decision making. It leads, for example, to overconfidence (hubris), motivated reasoning, loss aversion, poor risk assessment and substitution, that is, the tendency to replace the complex problem one is facing with a simpler one while pretending that they are in fact analogous (see e.g., Tversky and Kahneman

1973). Further, the stability of our judgements is affected by priming (Kiesel et al. 2007) and framing (Tversky and Kahneman 1981).

It seems reasonable to assume that these cognitive shortcomings hamper our general understanding of the world. We struggle to handle complex and contradictory information and this undermines our capacity for moral reasoning and our ability to make "good" decisions. By good decisions I mean socially responsible, reflective and rational decisions. Arguably, such decision making includes a propensity towards pro-social behavior, sustainability, sensitivity to long term consequences, willingness to assume an all-things-considered perspective, and a sense of equity. Failing to act responsibly, or virtuously, has very negative results not only for the individual but also on a societal level. Consider, for example, the recent harms linked to irresponsible individual and collective decisions about sustainable living, climate change, resource allocation, and public health.

While it might be clear that we ought to improve our thinking and decision making, it is of course an open question *exactly* which cognitive skills and virtues (moral and epistemic) we ought to prioritize in order to fare better both as individuals and as a collective. This chapter contains a few suggestions of capacities and virtues which, in combination, could have a positive impact on our belief forming, decision making and propensity to act in line with our moral values. The examples given here include improved working memory, higher cognitive flexibility, more focused and sustained attention, and increased capacity for learning. Such skills would facilitate the instilling of a set of virtues conducive to improved belief formation and good decision making in the broad sense, that is, with an eye to sustainability and fairness as well as the responsible development and use of technology.

4. Virtue Ethics and (Good) Decision Making

4.1 Thinking and Acting

Very broadly speaking, virtue ethicists tend to approach ethics by asking, "What is the good life?" For Aristotle the answer was that the supreme human good is *eudaimonia* (NE 1.1–1.13). This is the happy and fulfilled life, and it is rational to want it because it is only in this state of flourishing that one can exercise all one's capacities and be fully human. To live that life one needs to cultivate a set of moral and epistemic or intellectual virtues and then act in accordance with them. Examples of traditional moral virtues (i.e., those skills which make us excellent at *doing*) include courage, generosity and moderation. Examples of traditional epistemic or intellectual virtues (i.e., those skills which make us excellent at *thinking*) include: wisdom (*sophia*), intellect (*nous*) and scientific knowledge (*epistēmē*). The instilling of both types of virtues is done through a lengthy and often demanding education and habituation process and a lifelong commitment to the exercise of virtue.

For an agent to qualify as virtuous her actions must be performed consistently, for the sake of virtue and with pleasure. That is, a virtue has to be a habitual disposition and it is taken to give rise to relatively stable patterns of behavior (see e.g., NE Book 7; Burnyeat 1980; Sorabji 1980). The virtues will provide the motivation (i.e., desire), make us sensitive to the relevant (moral) factors in difficult situations, as well as enable us to deliberate well and reach the type of decisions which are conducive to our long-term well-being (on an individual and collective level). The mature moral decision-maker achieves this sensitivity through a combination of *phronesis (the epistemic/intellectual virtue of practical wisdom)* and the moral virtues, especially a sense of equity or justice. To have *phronesis* means to be good at thinking about how one should act in order to live a worthwhile life. Such an individual is good at thinking morally, that is, she knows the moral principles, she has a strong sense of equity and moreover, she knows how to apply them in practical situations. Notably then, to be virtuous means to be good at thinking morally but it is also about action—to be virtuous is to be habitually excellent at doing something.

However, even though morally mature decision making requires this type of holistic ability informed by all the virtues, one could imagine that some virtues would be particularly relevant for promoting good information handling. In addition to the above mentioned traditional Aristotelian virtues new—or non-traditional—virtues that would be especially useful might include open-mindedness, that is, the willingness to revise beliefs in the face of evidence and to entertain alternatives (Roberts and Wood 2007), epistemic conscientiousness/responsibility (Code 1984; Montmarquet 1987), intellectual honesty, fair-mindedness, tolerance, and impartiality. To see how virtues such as these might translate into behavior, consider how impartiality and intellectual honesty could mitigate the cognitive bias of motivated reasoning. I take it to be relatively uncontroversial that such virtues could contribute positively to how we form beliefs and to our well-being both now and in the future. The challenge is rather how to acquire them.

4.2 A Problem for the Virtue Ethicists

To be excellent is to have an unconditional disposition to act, feel and respond in ways typical of the good person. However, in light of the scientific findings listed in Section 3, it appears that many, if not most, of us do not have the cognitive skillset required for this type of excellence. This casts some doubt on two central themes in virtue ethics: first, that virtues are stable character traits issuing in action, and second, that these traits can be instilled through education, commitment, hard work, and training.

One might suspect, for example, that the recommended habituation is likely to produce mere "moral experts" and moral expertise that, although useful on occasion, is certainly not to be confused with "true excellence." Where the excellent person has a stable disposition and can be trusted to do the right thing, the moral expert is unreliable

(Schwitzgebel and Rust 2009). Sometimes she will act and respond as she ought, but on other occasions she will, for various reasons, not exercise the virtues. Admittedly, the negative impact of cognitive limitations on moral responsibility and decision making spell problems for all normative theories. However, it might be especially bad for virtue ethics since it has a focus on character building, deliberation, sensitivity to context and habituation and thus could be taken to be more cognitively and emotionally demanding than competing theories (see e.g., Swanton 2003). Might it be then that enhancement of core cognitive capacities is required to make the virtuous life a real possibility for most people?

5. Examples of Technologies that Can Be Used for Cognitive Enhancement

After confronting our cognitive shortcomings and their negative impact on our quality of life and general well-being, it might be a relief to learn that research in neuroscience and, especially, neurophysiology has also shown that the adult human brain has potential for structural and functional change (Watanabe et al. 1992; Zilles 1992; Stahnisch and Nitsch 2002; Pascual-Leone et al. 2011). The notion of *neuroplasticity* has generated much interest both inside and outside academia, but the fact that there is a potential for learning—and perhaps improvement—still leaves the question as to *how.* Through which means could positive, lasting and transferable cognitive changes with an acceptable balance between risk and potential benefit be achieved?

While this chapter will investigate a number of non-pharmacological, non-traditional technologies used for cognitive enhancement, I will very briefly mention some alternatives. Examples of pharmacological enhancers would be pharmaceutical drugs like methylphenidate, amphetamine, dopamine agonists, and modafinil but also hormones and neurotransmitters. Non-pharmaceutical traditional methods include education, physical exercise, diet, and supplements, as well as certain forms of mental training such as meditation. (Fröding and Osika 2015).

However, there is growing interest in non-pharmacological, non-traditional technologies that can be used for cognitive enhancement, just four of which I explore here: computer training, neurofeedback, tDCS (transcranial direct current technology, a form of non-invasive brain stimulation), and BCI (a form of computational neuroenhancement).[2]

While they each have therapeutic uses, I am interested in discussing the potential for improving core cognitive capacities in *healthy* humans, that is, people who already function within the normal range. Three key criteria of attractiveness of these technologies (setting aside for now the criterion of safety) include their (i) usefulness (i.e., might the practice plausibly improve or cultivate some cognitive capacity which in turn can facilitate instilling of virtues or hone our decision-making capacities?), (ii) durability (i.e., how long do the achieved effects last, and how soon do the effects begin to diminish?), and (iii) generalizability (i.e., is the skill transferable as opposed to task specific?).

5.1 Computer Training—Video Games and Apps

Much of the debate on computer games has focused on the widespread concern that gaming triggers anti-social behavior and aggression (Anderson et al. 2010; Greitemeyer and Mügge 2014) and undermines self-control (Gabbiadini et al. 2014). More recently, however, there has also been substantial discussion on whether or not playing certain kinds of computer games can have positive cognitive effects on, for example, working memory, the capacity for problem-solving, self-regulation, cognitive flexibility and attention.

Researchers have examined the positive effects of playing *action video games* on a range of perceptual, attentional and cognitive skills (Green and Bavelier 2015). This includes improvements in speed of processing (Dye et al. 2009a; Dye et al. 2009b), capacity to seek task-relevant information across space (Green and Bavelier 2003; Green and Bavelier 2006; Feng et al. 2007; Green and Bavelier 2007; Dye and Bavelier 2010; Wu and Spence 2013) and time (Li et al. 2010; Pohl et al. 2014), cognitive control and working memory (Colzato et al. 2013), ability to switch quickly between tasks (Karle et al. 2010; Cain et al. 2012; Colzato et al. 2014) and to carry out multiple tasks at the same time (Strobach et al. 2012; Chiappe et al. 2013). With regards to decision making, the action video game players seem better at identifying the information relevant to making accurate decisions (Green et al. 2010). Here, "accurate" means factually correct, not necessarily having a moral component.

While Shawn Green and Daphne Bavelier also include studies that have failed to show positive effects in their review, they conclude that

> While standard perceptual or cognitive training paradigms often produce learning that is highly specific to the exact context of the trained task, the benefits of action video game play have been shown to extend well beyond the confines of the games. Clear enhancements in basic perceptual skills, in the ability to utilize selective attention, and in cognitive flexibility have been noted as a result of action video game play (Green and Bavelier 2015, 106).

However, it should be noted that there is some disagreement as to how generalizable the skills actually are in practice (Simons et al. 2016; Lindenberger et al. 2017). Additional research shows that frequent playing of *pro-social video games* (i.e., games where cooperation is rewarded) can have a positive impact on behavior. For example, they may promote helpfulness, reduce negative cognitive-emotional constructs (e.g., stereotypes), promote some positive affective traits such as empathy (Gentile et al. 2009; Greitemeyer and Osswald 2010), and reduce aggression (Greitemeyer et al. 2012).

Yet other studies have looked at how playing *strategy video games* (characterized by long gaming sessions, planning and management of resources) can promote self-regulation. The researchers carried out two studies, one which was controlled for personality traits and individual preferences, and concluded that that frequent playing of such games is positively associated with self-regulation (Gabbiadini and Greitemeyer 2017). Plausibly, to be able to regulate one's emotions, resist impulses and act with an eye to the long-term best interest is central to the type of good decision making discussed in this chapter.

Within the domain of computer training there is of course also the ever growing industry of cognitive training apps, that is, apps that are designed to build cognitive capacities and promise various forms of cognitive improvements. Examples include Lumosity, Peak, BrainHQ and Elevate. The apps offer what is sometimes called "personalized brain training"; examples of skills users are said to sharpen include: memory, attention, flexibility, processing speed and problem solving. In addition, the users are said to stand to acquire further insights into cognition.

These commercial brain training apps are low risk, accessible and affordable. However, the scientific evidence regarding the usefulness, durability and generalizability of the training is not well documented. While there is no shortage of studies claiming proven results, there is substantial variation in research design and analysis which makes it hard to draw firm conclusions regarding the potential benefits of usage (Simons, Boot, Charness et al 2016). Indeed, in 2016 the US Federal Trade Commission fined the company Lumosity 2 million dollars for deceptive advertising and preying on consumers (Torous, Staples, Fenstermacher et al. 2016). A very large review study of the available peer-reviewed literature drew the following conclusion:

> We find extensive evidence that brain-training interventions improve performance on the trained tasks, less evidence that such interventions improve performance on closely related tasks, and little evidence that training enhances performance on distantly related tasks or that training improves everyday cognitive performance.
>
> (Simons, Boot, Charness et al 2016, 103)

Another recent study (Stojanoski, Lyons, Pearce et al. 2018) which primarily investigated transferability and generalizability of acquired skills came to the same conclusion: performance on training tasks improved, but that did not extend to test tasks.

5.1.2 *Methodological Problems*

Many studies of computer training games and apps use behavioral intervention which brings with it substantive methodological challenges. For example, the studies cannot be blinded and the users or players might be biased to be more motivated (Boot et al. 2011; Boot et al. 2013; Kristjánsson 2013; Green et al. 2014). There is an ongoing discussion about how to mitigate the effects on validity, for example, through using active control groups. A related problem is the difficulty of establishing a causal link (as opposed to mere association) between time spent playing and a certain cognitive skill. In addition, there has been a noticeable failure in replicating studies showing perceptual and cognitive differences between gamers and non-gamers. Examples include attempts to replicate studies on memory and executive control (Boot et al. 2008; Andringa and Boot 2017), visual information processing (van Ravenzwaaij et al. 2014), visual attention (Roque and Boot 2018), attention and memory (Cardoso-Leite et al. 2016), and dual task performance (Gaspar et al. 2014).

Hence, while there is a growing body of evidence indicating that action video games (the most studied type of game) can improve perceptual and cognitive skills, much more research is needed (Gentry et al. 2019). For one thing, larger samples are essential to learn more about the impact of individual differences on study results. Also, it is necessary to further explore the potential for skill transfer in order to learn more about if and how the improved cognitive skills could inform everyday decisions and behavior. Further, it should be noted that for any lasting positive effect on cognitive capacities the games need to be played frequently (usually 30-50h), for longer periods of time and at regular intervals.

Notably, it is becoming increasingly common to combine gaming and neurofeedback. Consider for example EEG gaming headsets designed to improve attention and focus, or VR glasses (and indeed whole suits) designed to enhance the game experience. While many, if not most, such neuro-prosthetic devices would require significant development (functionality, design and user-friendliness) to ensure wide uptake, it is not unlikely that advanced versions of such devices will become a part of many people's lives and indeed something which we take for granted. We will now turn to discuss some aspects of neurofeedback.

5.2 Neurofeedback (or EEG Biofeedback)

Neurofeedback (NFB) is a specific form of biofeedback and its purpose is to teach the trainee to exert "control over specific EEG parameters and thus to influence associated cognitive functions" (Dessy et al. 2018, 34). In practice, the trainee undergoes a series of training sessions and gradually—through trial and error—becomes able to "modify the brain activity and learn to self-regulate his or her EEG activity" (Dessy et al. 2018, 15).

When the trainee manages to produce the intended changes in the brainwave pattern there will be a reward in the form of an auditory or visual signal. For a very concrete example, imagine that you are watching a movie. After a while your mind starts wandering and you lose focus. This change in your brainwave pattern will cause the picture to blur—this is the penalty—you refocus and the picture gets sharp—the brain is rewarded. Note that an ongoing feedback is essential to the learning process, as it enables the desired neurophysiological changes to consolidate (Evans and Abarbanel 1999).

EEG NFB is commonly used in clinical settings to treat a range of neurobiological dysfunctions; for example, ADHD, autism spectrum disorders, substance use, epilepsy and learning difficulties (Niv 2013). As shown in a recent review article (Dessy et al. 2018), however, there is a growing body of research looking at different NFB training protocols as a method for enhancing cognitive performance in a non-clinical population. Examples of reported positive cognitive effects include: memory consolidation

(Reiner et al. 2014; Rozengurt et al. 2016), improvements in executive functions in young subjects (Wang and Hsieh 2013), improved short-term memory frequency (Nan et al 2012; Hsueh et al. 2016), improved semantic processing in a working memory task (Vernon et al. 2003), enhanced attentional performance (Fritson et al. 2008; Doppelmayr and Weber 2011), reduced psycho-emotional stress (Bazanova et al. 2013), and improved declarative memory performance (Hoedlmoser et al. 2008), as well as positive effects on familiarity-based processes in working memory (Guez et al. 2015).

As pointed out by Dessy et al. the possibility of achieving lasting effects seems to require a certain amount of virtue. For the desired biological change to come about—that is, to achieve the amount of over-learning required for automatization—the trainee needs to be committed and self-disciplined. Just as importantly she needs to know when to stop to avoid negative results (Dessy et al. 2018, 38). In other words she would require both individual moral and epistemic virtues like moderation and practical wisdom to inform an understanding of what type of education and activities she ought to engage in, all things considered.

Currently there is little or no literature stating the exact number of sessions required. Indeed, large variations between individuals regarding the responsivity to EEG NFB have been reported (Enriquez-Geppert et al. 2017, 10). Further, there is a lack of consistency in methodology, terminology, training protocol and frequency range selection (Dessy et al. 2018, 38) and we know little about the long-term effects. Clearly, much more research is needed.

That said, it does not seem implausible that normally functioning individuals could, through a combination of computer training (cultivating both cognitive and emotional skills) and EEG NFB, strengthen some cognitive capacities and decrease some disabling emotions (e.g., anxiety, misplaced fear), in ways that would be conducive to good decision making. Virtue ethics could provide some guidance as to the prudent development of such technologies as well as how to use them to train in a way that is conducive to the good life.

5.3 Non-invasive Brain Stimulation—tDCS

Transcranial direct stimulation (tDCS) is an example of a non-invasive brain stimulation technology which is considered to have potential for cognitive enhancement. tDCS works by manipulating the neurons with a weak electric current that causes changes in excitation and inhibition. The current is delivered through electrodes which are placed on the scalp of the participant. The position of the electrodes depends on which brain region is to be studied and the current is delivered for 5-20 minutes depending on protocol. The effects of this non-invasive modulation are dependent on precision of the stimulation, so the most successful examples of tDCS tend to be fMRI-guided.

A number of studies have shown that tDCS can have enhancing effects on spatial learning and memory (De Jongh et al. 2008; Chi et al. 2010; Hamilton, Messing, and Chatterjee 2011; Kadosh et al. 2012). Further, tDCS has been shown to reduce false memories (Boggio et al. 2009; Gallate et al. 2009), increase verbal fluency (Iyer et al. 2005) and even promote a more careful driving style in a car simulation experiment. As pointed out by Dresler et al. however, the effects on driving style might be a consequence of reduced risk-taking on the driver's part, rather than an improvement of a more general skill like planning (Dresler et al. 2013, 536).

More recently, researchers have become interested in using tDCS to enhance social behavior in healthy individuals. One such study (Sellaro et al. 2015) wanted to examine how the medial prefrontal cortex (mPFC) may contribute to self-regulatory and cognitive-control processes implemented to override social stereotypes. The researchers used tDCS and concluded that their results "provide evidence for a critical role of the mPFC in counteracting stereotypes activation. Furthermore, our results are consistent with previous findings showing that increasing cognitive control may overcome negative bias toward members of social out-groups" (Sellaro et al. 2015, 891). Additional studies have looked at how tDCS can be used to promote other aspects of self-other representations which are crucial to successful social interaction, for example, by testing performance on perspective-taking tasks and control-of-imitation tasks.

As with the other technologies introduced in this chapter, tDCS shows substantial variation between individuals, reported cognitive effects tend to be modest, there are questions as to the durability of these effects, and there is no evidence of *generally* enhancing cognitive effects. In addition, there is the criticism that the vast majority of the experiments are about testing specific tasks in a controlled laboratory, and that it is not clear that the same positive effects on performance would manifest themselves in a real-life setting. While non-invasive brain stimulation technologies like tDCS are not considered high-risk compared to, for example, deep brain stimulation, there is a risk of premature use as a consequence of technology hype or speculation (Dresler et al. 2013) and the long-term effects are largely unknown (Kadosh et al. 2012).

5.4 Brain-Computer Interfaces

BCI is a technology which creates a connection between a human brain and a computer. This allows the brain to communicate directly with the computer (Mak and Wolpaw 2009; van Gerven et al. 2009). BCIs can be non-invasive (EEG based), partially invasive (devices are implanted inside the skull but rest outside the brain rather than within the grey matter), or invasive. Invasive BCI means that the electrodes are implanted in the grey matter of the brain. This provides a very good signal (something which is a challenge for non-invasive versions) but involves complex surgery and is thus high-risk. In addition, the technology is not wireless and it is very expensive.

Currently BCIs are mostly one-directional (i.e., not involving neurofeedback) and used to enable people who suffer brain injuries and paralysis to move and to communicate. (Birbaumer and Cohen 2007; Birbaumer et al. 2008; Shih et al. 2012). The BCI converts the person's intent into action; for example, the technology may allow the patient to control their limbs with their thoughts. BCIs are also used to aid people suffering from Parkinson's (Little et al. 2013), and there is a discussion on the usefulness of BCIs for assessment and treatment of psychopathy (Jotterand and Giordano 2015), as well as ongoing research on the encoding and recalling of memories (Berger et al. 2011; Song et al. 2016). As for the memory implantation, however, the effects vary greatly. As pointed out by Burke et al., "Understanding and reducing this variability represents the main hurdle in the realization of a mnemonic BCI to enhance memory formation, and should be the focus of future research" (Burke et al. 2015).

Despite the challenges many see great potential for BCI cognitive enhancement. A popular view is that this type of technology will enable us to "decode people's mental processes and directly manipulate the brain mechanisms underlying their intentions, emotions and decisions; where individuals could communicate with others simply by thinking; and where powerful computational systems linked directly to people's brains aid their interactions with the world such that their mental and physical abilities are greatly enhanced" (Yuste et al. 2017, 160).

Two examples of BCI projects researching what Dresler et al. refer to as the "joint outputs of minds coupled with machines" (Dresler et. al, 2019, 1139) are Facebook and Neuralink (the latter is owned by Elon Musk). The latter company aims to develop a type of "neural lace" which connects the brain to a computer and involves neurofeedback. Neural lace would differ from most of the BCI currently used for therapeutic purposes. Neural lace is a mesh that will be inserted via a needle into the head and then unravel itself and become a layer on top of the brain. The lace would, in theory, become a part of the brain and function like an interface allowing for the human brain to interact with computers wirelessly. The fact that it enables neurofeedback means that it would be (at least in theory) possible to enhance learning and behavior change, in an unprecedented way. Musk described his vision as a "merger between biological intelligence and machine intelligence" (Solon 2017). Another branch of industry which takes great interest in BCIs would be defense. From a military perspective, there is great potential for BCI both on and off the battlefield; for example, to restore neural and behavioral function in soldiers, accelerate learning, and improve threat detection abilities (Miranda et al. 2015).

While there is much speculation and hype around BCIs, there is very little in the way of published results indicating actual enhancement effects in normally functioning humans. That said, while the technology is further off in the future than the other examples in this chapter, it is—as observed by many experts—a real possibility. Consequently, there is a need for an informed, transparent and inclusive discussion on the ethical aspects of developing, distributing and regulating this and other similar technologies. Key ethical issues include privacy, autonomy, agency and identity, reinforcement of bias, dual use, diffusion of responsibility as well as concerns regarding justice, equity and diversity (Lucivero and Tamburrini 2008; Clausen 2011; Vlek et al. 2012; O'Brolcháin and Gordijn 2015; Yuste et al. 2017). Arguably, virtue ethics (especially in combination with other cognitive enhancement strategies such as tailored education, see concluding section) could inform this deliberation, make us better prepared and increase the chances that the technology is used responsibly and for the good of society.

6. Virtue and Technology

As shown above, there appears to be some evidence of potential enhancing effects of these new technologies on core cognitive skills in normally functioning humans, but much more research on their transferability, durability, individual variation, and long-term effects is required. So let us turn to the question of why a combination of technology and virtue would be attractive.

It *might* be the case that enhancing cognitive flexibility as well as focused and sustained attention could reduce bias and experiential avoidance, and improve emotional regulation, which in turn could promote more impartial and pro-social behavior. It could even be the case that cognitive enhancement could deepen the understanding of why the virtuous life is the good life, and strengthen the commitment to the pursuit of such a life. To simply assume that good moral decision making automatically would be developed in tandem with such cognitive improvements appears, however, to be a risky strategy. Indeed, an often-voiced concern about cognitive enhancement is that it would bring about enhanced individuals who might not allow the relevant moral considerations to inform their deliberation. In other words—cognitively enhanced individuals might be clever and *vicious*.

Hence, a more prudent approach would be to embed and anchor the cognitive capacities in a virtue framework which would facilitate the development of *moral* reasoning skills and an overall understanding of how one ought to live.[3] On a general level, to be virtuous means having sound judgement, a sense of equity and an ability to take an all-things-considered perspective. For more concrete examples, consider the virtues listed earlier: intellectual honesty, *phronesis*, open-mindedness, tolerance, impartiality, fair-mindedness, capacity for introspection, epistemic conscientiousness, and a sense of reciprocity. Presumably, such capacities are highly conducive to responsible decision making. They could, for example, inform our risk-assessment and balance some of the systemic bias and motivated reasoning most of us are guilty of.

Next I sketch two ways in which virtue ethics and cognitive enhancement technologies are mutually supportive. First, a virtue framework could play a big role in promoting the responsible use of these and other technologies. Second, the successful instilling of the virtues might (for many people) require cognitive enhancement.

6.1 Facilitating Responsible Technology Decisions

As Aristotle pointed out, ethics is not a science and cannot be codified as a set of rules (*Nicomachean Ethics*: see NE I.3, 1094b11–27; I.7, 1098a26–34), and even if it could, the very nature of emerging technologies—that is, they develop so quickly that lawmakers and regulators have a hard time keeping up—threatens to make the rules obsolete. Even worse, reliance on a rule-based approach could create a sense of false security and rigidity in thinking about the development, use, and regulation of enhancement technologies. It appears that what is required is the cultivation of an overall sensitivity and ability to identify morally relevant features of such technologies and then be moved to action guided by the moral and epistemic virtues. To bring home this last point, consider, for example, how irresponsible handling of synthetic biology, nanotechnology, AI, and machine learning could not only diminish our quality of life but actually undermine our very existence. We need to get better at identifying (including foreseeing), preventing, and, failing that, at least mitigating the risks attached to emerging technologies, including those used for cognitive enhancement. Virtue is thus a prerequisite for responsible use and further development of cognitive enhancement technology (Vallor 2016).

Further to the desirability of a combination of cognitive enhancement and virtue, it would seem that the virtues would add the necessary commitment and self-discipline required for any lasting effects of, for example, computer gaming and neurofeedback practice. The virtues would improve our understanding not only of which type of cognitive enhancement technologies we ought to pursue, but also how much we should train and when we should stop (to avoid negative results).

6.2 Making the Good Life a Real Possibility

A common argument against enhancement is that it is cheating, and that the enhanced gain provides an unfair advantage over one's peers. This concern is mostly voiced in the context of physical enhancement such as doping in sports (Schermer 2008), but it might also be levelled at cognitive enhancement from a virtue perspective. I imagine that the worry is roughly this: there is an element of hard work that is central to virtue ethics. The virtues, including epistemic virtues, should be instilled through education and habituation and be fine-tuned over a lifetime; the fact that this takes time and dedication is important. While enhanced individuals might not gain an unfair advantage as such, they would cheat themselves and miss out not only on an important process but also on a dimension of virtue that is only available if one goes through that type of instilling.

I quite agree—the process of habituation is key to instilling the virtues and shapes us in a way that enables us to be sensitive to the relevant features in a situation, to take pleasure in the right things, for the right reasons, to the right extent, and so on. This process could indeed be said to have intrinsic value and it is hard to see how cognitive enhancements of the sort discussed here could mimic all the worthwhile aspects of that process.

However, it seems highly unlikely that the type of cognitive enhancements that have been discussed here would make virtue "too easy"—they are far too piecemeal. Even assuming that some of the effects described in this chapter could be generalized outside the specific training situation and were durable, they would not furnish us with the holistic skillset required for good decision making. To make the type of all-things-considered decisions that, arguably, would be required for ethical handling of technology or any other important concern, we need to be informed by a range of epistemic and moral virtues. Consequently, for most people, character building would remain a lifelong undertaking which would require great effort and commitment. The moral and epistemic virtues would still have to be acquired in stages, through education and habituation and they have both cognitive and emotional dimensions. That is, the virtuous agent must desire and take pleasure in doing the right thing, it is not enough to simply make an intellectual choice to do X rather than Y.[4]

Building on the above "no replacement argument," I suggest that the combination of cognitive enhancement technology and virtues could go some way to mitigate a frequent criticism of virtue ethics, namely that it is so demanding that it is effectively impossible (Fröding 2011). This critique is often based on the idea that the way people act has more to do with the particular situation or their circumstances and less with stable character traits (see e.g., Nussbaum 2001; Hursthouse 1991; Harman 1999; Doris 2002; Haidt et al. 2008).

The role of the cognitive enhancement would be to help create a more stable platform for the individual to start off from. By reducing some of the cognitive weaknesses, such as bias and experiential avoidance, and improving other capacities, such as focused and sustained attention and cognitive flexibility, individuals would be levelled up to a starting point where more people could—through great effort—develop the virtues that will guide and inform their decisions and actions (Fröding 2011, Fröding and Osika 2015).

It is likely that improved cognitive and emotional regulation skills, coupled with epistemic and moral virtues, could make us more stable, able, and more motivated to continue to cultivate the virtues as we improve our understanding of what the good and happy life is. The commitment to, and practice of, virtue is in and of itself an embedding strategy, and increased ability for habituation (and deeper understanding of what the good life is) could plausibly have a positive impact on motivation (see e.g. Niemiec and Ryan 2009; Deci and Ryan 2012). Some cognitive enhancements and a commitment to virtue could work as a virtuous loop where you get increasingly motivated to further your ability (Fröding and Osika 2015).

The improved cognitive and emotional regulation capacities would aid in the development of emotional regulation and self-detachment (in the sense of "considering others") and by decreasing experiential avoidance it would make us more able to handle the internal and external conflicts which might impede good decision making (Richter et al. 2019). Consider for example, the study by Gabbiadini and Greitemeyer (Section 5.1) where it was concluded that frequent playing of strategy video games is positively associated with self-regulation. Improved self-regulation and improved ability for resisting impulses would contribute to the type of good decision making explored in this chapter. It should, however, be noted that this was a correlation study and thus one should be cautious about causality.

Another example would be the study by Green and Bevalier (Section 5.1) where improved cognitive flexibility was noted as a result of action video game play. Whilst the generalizability has been questioned, increased cognitive flexibility would facilitate the instilling of a set of virtues which are conducive to good belief forming and good decision making (open mindedness, epistemic conscientiousness, intellectual honesty, fair mindedness, tolerance and impartiality). Similarly, another core cognitive capacity—focused and sustained attention—has been shown to improve in adolescents through certain types of video games (Patsenko et al. 2019).

As previously pointed out, much more research (including better tailored methods) is required for us to learn how, and to what extent, video games can shape cognitive performance (Dale et al. 2020). Looking further ahead, BCIs like neural lace technology (see Section 5.4) with its neurofeedback has the potential to radically change the way we learn and how we make decisions. These too could plausibly impact our propensity and capacity for developing epistemic and moral virtues conducive to good decision making.

7. Conclusion: Ethical and Practical Challenges

In addition to being safe, we want cognitive enhancements to be beneficial and reliable. It is clear that none of the technologies exemplified here tick all those boxes yet. Little is known of the long-term consequences, it is not clear that the skills gained will be useful in wide application, and there are questions about the duration of enhancing effects, to mention but a few problems.

A common concern focuses on the piecemeal nature of the proposed enhancements. Many cognitive abilities (e.g. creativity, complexity awareness and complexity management skills) are multi-factorial and it is unlikely that we, in the near future, will be able to improve them in a radical, comprehensive transformation. Another worry is that many enhancements could involve trade-offs. For example, increased pro-social behavior might come at the cost of survival skills which, if true, would undermine the attractiveness of the enhancement (see e.g. Shickle 2000). Further, there is a host of other ethical concerns that need to be taken into account and properly addressed if we are to develop and use enhancement technology in a responsible way, most notably safety, autonomy, privacy, agency, justice, fairness and long-term consequences.

Equally, it is widely recognized that these and other technologies have great potential and the complex collective action problems our species is facing, e.g. how to achieve social and environmental sustainability or manage global health threats, are of a highly pressing nature. Without over-extending, or pretending that such issues will be easily solved, I do think that a combination of cognitive enhancement and virtue is a promising strategy. It is evident that we need to get more cognitively skilled, but for good decision making we also need the epistemic and moral virtues. While cognitive improvements by means of technology can facilitate instilling the virtues and thus be conducive to (possibly even necessary for) leading the good life, they are not in and of themselves enough. I believe this holds true even if we assume a significantly more positive effect on cognition than what has been documented so far. What the virtues contribute is an awareness of the morally relevant parameters in a situation, and they transform the agent into a stable and reliable decision maker. They harness the prerequisite cognitive skills and provide the robust yet flexible moral framework which will be required for leading the good life in an increasingly complex society. Equipped with such skills for deliberation and a sense of equity, the agent is more likely to both get the information right and to proceed to make the right decision in collaboration with other agents.

However, since this technology is in its very early stages, it would be prudent to engage in a combination of strategies for cognitive improvement. While this chapter has been about enhancement through technology, the role of traditional education should not be underestimated (see e.g. Kristjánsson 2015). Education is a powerful tool and can be combined with various technologies as well as other lifestyle choices as broadly conceived of (e.g., physical exercise, meditation, diet) to boost the cognitive effect. Such strategies are not mutually exclusive; to the contrary they often make for good combinations, e.g. tailored education packages including neurofeedback and computer training in combination with traditional education and character education for virtue. When combined, these methods can help increase commitment to and understanding of the moral and epistemic virtues and why the good life is the best life (Fröding and Osika 2015). Such a combined strategy would simultaneously improve the capacity for responsible development and handling of emerging technologies (for enhancement and other purposes).

A further contributing factor would be a social structure which encourages and rewards virtuous behavior (e.g., collaboration, fairness, tolerance). This would, for example, include an inclusive, society-wide, ongoing dialogue both on moral and epistemic values and on how they might be exercised, that is, translated into actual behavior.

Some scholars are concerned that cognitive enhancement—not only the radical kind but also the more moderate versions discussed here—would undermine our appreciation of life and pose a threat to human bonding (e.g., Sandel 2009). I would argue the opposite. Some cognitive enhancement coupled with virtue ethics could make us, if not more human, at least more *humane*: more responsible, more likely to extend our care and concern, more committed to fairness and reciprocity (including increased conflict resolution capacities, e.g., Klimecki 2019), more able to make long term sustainable choices and more likely to make balanced and informed decisions with regards to the development and use of technology.

Notes

1. I have defended these ideas elsewhere, and parts of this chapter draw on previously published research, e.g., Fröding and Osika 2015; Fröding 2011; Fröding 2012.
2. Here I follow Bert Gordjin's broad definition of neuroenhancement: "an intervention in the central nervous system, by using pharmaceutical means, surgery, and/or technology

(brain-computer interfaces or other neurotechnologies), in order to "improve" certain aspects of its "healthy" or "normal" performance." (Gordjin 2015, 1171).

3. For the argument that specific moral enhancement is required, see Persson and Savulescu 2012; Douglas 2008.
4. Notably the relationship between the intellectual and the emotional in moral decision making is among the most disputed issues in the *Nicomachean Ethics*. See NE Book 3.2–3.3 and NE Book 6.12–16.13.

References

Anderson, Craig A., Akiko Shibuya, Nobuko Ihori, Edward L. Swing, Brad J. Bushman, Akira Sakamoto, Hannah R. Rothstein, and Muniba Saleem. 2010. "Violent Video Game Effects on Aggression, Empathy, and Prosocial Behavior in Eastern and Western Countries: A Meta-analytic Review." *Psychological Bulletin* 136, no. 2: 151.

Andringa, Ronald, and Walter R. Boot. 2017. "Video Games." In *Theory-Driven Approaches to Cognitive Enhancement*, 199–210. Springer, Cham.

Arnsten, Amy FT. 2009. "Stress Signalling Pathways that Impair Prefrontal Cortex Structure and Function." *Nature Reviews Neuroscience* 10, no. 6: 410.

Bazanova, O. M., N. V. Balioz, K. B. Muravleva, and M. V. Skoraya. 2013. "Effect of Voluntary EEG α Power Increase Training on Heart Rate Variability." *Human Physiology* 39, no. 1: 86–97.

Berger, Theodore W., Robert E. Hampson, Dong Song, Anushka Goonawardena, Vasilis Z. Marmarelis, and Sam A. Deadwyler. 2011. "A Cortical Neural Prosthesis for Restoring and Enhancing Memory." *Journal of Neural Engineering* 8, no. 4: 046017.

Birbaumer, Niels, and Leonardo G. Cohen. 2007. "Brain–Computer Interfaces: Communication and restoration of Movement in Paralysis." *The Journal of physiology* 579, no. 3: 621–636.

Birbaumer, Niels, Ander Ramos Murguialday, and Leonardo Cohen. 2008. "Brain–Computer Interface in Paralysis." *Current Opinion in Neurology* 21, no. 6: 634–638.

Boggio, Paulo S., Soroush Zaghi, and Felipe Fregni. 2009. "Modulation of Emotions Associated with Images of Human Pain Using Anodal Transcranial Direct Current Stimulation (tDCS)." *Neuropsychologia* 47, no. 1: 212–217.

Boot, Walter R., Arthur F. Kramer, Daniel J. Simons, Monica Fabiani, and Gabriele Gratton. 2008. "The Effects of Video Game Playing on Attention, Memory, and Executive Control." *Acta Psychologica* 129, no. 3: 387–398.

Boot, Walter R., Daniel J. Simons, Cary Stothart, and Cassie Stutts. 2013. "The Pervasive Problem with Placebos in Psychology: Why Active Control Groups Are not Sufficient to Rule out Placebo Effects." *Perspectives on Psychological Science* 8, no. 4: 445–454.

Boot, Walter Richard, Daniel P. Blakely, and Daniel J. Simons. 2011. "Do Action Video Games Improve Perception and Cognition?" *Frontiers in Psychology* 2: 226.

Bostrom, Nick, and Rebecca Roache. 2008. "Ethical Issues in Human Enhancement." In *New Waves in Applied Ethics*, edited by Jesper Ryberg, Thomas Petersen, and Clark Wolf, 120–152. London: Palgrave Macmillan.

Bostrom, Nick, and Anders Sandberg. 2009. "Cognitive Enhancement: Methods, Ethics, Regulatory Challenges." *Science and Engineering Ethics* 15, no. 3: 311–341.

Burke, John F., Maxwell B. Merkow, Joshua Jacobs, Michael J. Kahana, and Kareem A. Zaghloul. 2015. "Brain Computer Interface to Enhance Episodic Memory in Human Participants." *Frontiers in Human Neuroscience* 8: 1055.

Burnyeat Michael. 1980. "Aristotle on Learning to be Good." In *Essays On Aristotle's Ethics*, edited by Amelia O. Rorty, 69–92. University of California Press.

Cain, Matthew S., Ayelet N. Landau, and Arthur P. Shimamura. 2012. "Action Video Game Experience Reduces the Cost of Switching Tasks." *Attention, Perception, & Psychophysics* 74, no. 4: 641–647.

Cardoso-Leite, Pedro, Rachel Kludt, Gianluca Vignola, Wei Ji Ma, C. Shawn Green, and Daphne Bavelier. 2016. "Technology Consumption and Cognitive Control: Contrasting Action Video Game Experience with Media Multitasking." *Attention, Perception, & Psychophysics* 78, no. 1: 218–241.

Chi, Richard P., Felipe Fregni, and Allan W. Snyder. 2010. "Visual Memory Improved by Non-invasive Brain Stimulation." *Brain Research* 1353: 168–175.

Chiappe, Dan, Mark Conger, Janet Liao, J. Lynn Caldwell, and Kim-Phuong L. Vu. 2013. "Improving Multi-tasking Ability through Action Videogames." *Applied Ergonomics* 44, no. 2: 278–284.

Clausen, Jens. 2011. "Conceptual and Ethical Issues with Brain–Hardware Interfaces." *Current Opinion in Psychiatry* 24, no. 6: 495–501.

Code, Lorraine. 1984. "Toward a Responsibilist Epistemology." *Philosophy and Phenomenological Research* 45, no. 1: 29–50.

Colzato, Lorenza S., Wery P. M. van den Wildenberg, and Bernhard Hommel. 2014. "Cognitive Control and the COMT Val 158 Met Polymorphism: Genetic Modulation of Videogame Training and Transfer to Task-switching Efficiency." *Psychological research* 78, no. 5: 670–678.

Colzato, Lorenza S., Wery P. M. van den Wildenberg, Sharon Zmigrod, and Bernhard Hommel. 2013. "Action Video Gaming and Cognitive Control: Playing First Person Shooter Games Is Associated with Improvement in Working Memory But Not Action Inhibition." *Psychological Research* 77, no. 2: 234–239.

Dale, Gillian, et al. 2020. "A New Look at the Cognitive Neuroscience of Video Game Play." *Annals of the New York Academy of Sciences* 1464.1: 192–203.

Daniels, Norman. 2007. *Just Health: Meeting Health Needs Fairly*. Cambridge University Press.

De Jongh, Reinoud, Ineke Bolt, Maartje Schermer, and Berend Olivier. 2008. "Botox for the Brain: Enhancement of Cognition, Mood and Pro-Social Behavior and Blunting of Unwanted Memories." *Neuroscience & Biobehavioral Reviews* 32, no. 4: 760–776.

Deci, Edward L., and Richard M. Ryan. 2012. "Self-Determination Theory." In *Handbook of Theories of Social Psychology,* edited by Paul Van Lange, Arie Kruglanski, and E. Tory Higgins, 416–436. Sage Publications Ltd.

Dessy, Emilie, Martine Van Puyvelde, Olivier Mairesse, Xavier Neyt, and Nathalie Pattyn. 2018. "Cognitive Performance Enhancement: Do Biofeedback and Neurofeedback Work?." *Journal of Cognitive Enhancement* 2: 12–42.

Doppelmayr, Michael, and Emily Weber. 2011. "Effects of SMR and theta/beta Neurofeedback on Reaction Times, Spatial Abilities, and Creativity." *Journal of Neurotherapy* 15, no. 2: 115–129.

Doris, John. 2002. *Lack of character: Personality and moral behavior*. Cambridge: Cambridge University Press.

Douglas, Thomas. 2008. "Moral enhancement." *Journal of Applied Philosophy* 25, no. 3: 228–245.

Dresler, Martin, Anders Sandberg, Christoph Bublitz, Kathrin Ohla, Carlos Trenado, Aleksandra Mroczko-Wąsowicz, S. Kühn, and Dimitris Repantis. 2019. "Hacking the Brain: Dimensions of Cognitive Enhancement." *ACS Chemical Neuroscience* 10, no. 3.

Dresler, Martin, Anders Sandberg, Kathrin Ohla, Christoph Bublitz, Carlos Trenado, Aleksandra Mroczko-Wąsowicz, Simone Kühn, and Dimitris Repantis. 2013. "Non-pharmacological Cognitive Enhancement." *Neuropharmacology* 64: 529–543.

Dye, Matthew W. G., and Daphne Bavelier. 2010. "Differential Development of Visual Attention Skills in School-Age Children." *Vision Research* 50, no. 4: 452–459.

Dye, Matthew W. G., C. Shawn Green, and Daphne Bavelier. 2009A. "Increasing Speed of Processing with Action Video Games." *Current Directions in Psychological Science* 18, no. 6: 321–326.

Dye, Matthew W. G., C. Shawn Green, and Daphne Bavelier. 2009B. "The Development of Attention Skills in Action Video Game Players." *Neuropsychologia* 47, no. 8–9: 1780–1789.

Enriquez-Geppert, Stefanie, et al. 2017. "Cognitive Enhancement by Self-Regulation of Endogenous Oscillations with Neurofeedback." In *Theory-Driven Approaches to Cognitive Enhancement* (1st ed.), edited by Lorenza Colzato. New York: Springer.

Evans, James R., and Andrew Abarbanel, eds. 1999. *Introduction to Quantitative EEG and Neurofeedback*. Amsterdam: Elsevier.

Feng, Jing, Ian Spence, and Jay Pratt.2007. "Playing an Action Video Game Reduces Gender Differences in Spatial Cognition." *Psychological Science* 18, no. 10: 850–855.

Fritson, Krista K., Theresa A. Wadkins, Pat Gerdes, and David Hof. 2008. "The Impact of Neurotherapy on College Students' Cognitive Abilities and Emotions." *Journal of Neurotherapy* 11, no. 4: 1–9.

Fröding, Barbro. 2011. "Cognitive Enhancement, Virtue Ethics and The Good Life." *Neuroethics* 4, no. 3: 223–234.

Fröding, Barbro. 2012. *Virtue Ethics and Human Enhancement*. Springer Science & Business Media.

Fröding, Barbro, and Walter Osika. 2015. *Neuroenhancement: How Mental Training and Meditation Can Promote Epistemic Virtue*. New York, NY: Springer International Publishing.

Gabbiadini, Alessandro, and Tobias Greitemeyer. 2017. "Uncovering the Association between Strategy Video Games and Self-Regulation: A Correlational Study." *Personality and Individual Differences* 104: 129–136.

Gabbiadini, Alessandro, Paolo Riva, Luca Andrighetto, Chiara Volpato, and Brad J. Bushman. 2014. "Interactive Effect of Moral Disengagement and Violent Video Games on Self-Control, Cheating, and Aggression." *Social Psychological and Personality Science* 5, no. 4: 451–458.

Gallate, Jason, Richard Chi, Sophie Ellwood, and Allan Snyder. 2009. "Reducing False Memories by Magnetic Pulse Stimulation." *Neuroscience Letters* 449, no. 3: 151–154.

Gaspar, John G., Mark B. Neider, James A. Crowell, Aubrey Lutz, Henry Kaczmarski, and Arthur F. Kramer. 2014. "Are Gamers Better Crossers? An Examination of Action Video Game Experience and Dual Task Effects in a Simulated Street Crossing Task." *Human Factors* 56, no. 3: 443–452.

Gentile, Douglas A., Craig A. Anderson, Shintaro Yukawa, Nobuko Ihori, Muniba Saleem, Lim Kam Ming, Akiko Shibuya et al. 2009. "The Effects of Prosocial Video Games on Prosocial Behaviors: International Evidence from Correlational, Longitudinal, and Experimental Studies." *Personality and Social Psychology Bulletin* 35, no. 6: 752–763.

Gentry, Sarah Victoria, Andrea Gauthier, Beatrice L'Estrade Ehrstrom, David Wortley, Anneliese Lilienthal, Lorainne Tudor Car, Shoko Dauwels-Okutsu et al. 2019. "Serious

Gaming and Gamification Education in Health Professions: Systematic Review." *Journal of Medical Internet Research* 21, no. 3: e12994.

Golkar, Armita, Emilia Johansson, Maki Kasahara, Walter Osika, Aleksander Perski, and Ivanka Savic. 2014. "The Influence of Work-Related Chronic Stress on the Regulation of Emotion and on Functional Connectivity in the Brain." *PLoS One* 9, no. 9: e104550.

Gordjin, Bert. 2015. "Neuroenhancement." In *Handbook of Neuroethics,* edited by Jens Clausen and Neil Levy, 1169–1175. Springer, Dordrecht.

Green, C. Shawn, and Daphne Bavelier. 2003. "Action Video Game Modifies Visual Selective Attention." *Nature* 423, no. 6939: 534–537.

Green, C. Shawn, and Daphne Bavelier. 2006. "Effect of Action Video Games on the Spatial Distribution of Visuospatial Attention." *Journal of Experimental Psychology: Human Perception and Performance* 32, no. 6: 1465–1478.

Green, C. Shawn, and Daphne Bavelier. 2007. "Action-Video-Game Experience Alters the Spatial Resolution Of Vision." *Psychological science* 18, no. 1: 88–94.

Green, C. Shawn, and Daphne Bavelier. 2015. "Action Video Game Training for Cognitive Enhancement." *Current Opinion in Behavioral Sciences* 4: 103–108.

Green, C. Shawn, Alexandre Pouget, and Daphne Bavelier. 2010. "Improved Probabilistic Inference as a General Learning Mechanism with Action Video Games." *Current Biology* 20, no. 17: 1573–1579.

Green C. Shawn, Tilo Strobach, and Torsten Schubert. 2014. "On Methodological Standards in Training and Transfer Experiments." *Psychological Research* 78:756–772.

Greitemeyer, Tobias, Maria Agthe, Robin Turner, and Christina Gschwendtner. 2012. "Acting Prosocially Reduces Retaliation: Effects of Prosocial Video Games on Aggressive Behavior." *European Journal of Social Psychology* 42, no. 2: 235–242.

Greitemeyer, Tobias, and Dirk O. Mügge. 2014. "Video Games Do Affect Social Outcomes: A Meta-Analytic Review of the Effects of Violent and Prosocial Video Game Play." *Personality and Social Psychology Bulletin* 40, no. 5: 578–589.

Greitemeyer, Tobias, and Silvia Osswald. 2010. "Effects of Prosocial Video Games on Prosocial Behavior." *Journal of Personality and Social Psychology* 98, no. 2: 211–221.

Guez, Jonathan, Ainat Rogel, Nir Getter, Eldad Keha, Tzlil Cohen, Tali Amor, Shirley Gordon, Nachshon Meiran, and Doron Todder. 2015. "Influence of Electroencephalography Neurofeedback Training on Episodic Memory: A Randomized, Sham-Controlled, Double-Blind Study." *Memory* 23, no. 5: 683–694.

Haidt, Jonathan, J. Patrick Seder, and Selin Kesebir. 2008. "Hive Psychology, Happiness, and Public Policy." *Journal of Legal Studies* 37, no. S2: S133–S156.

Hamilton, Roy, Samuel Messing, and Anjan Chatterjee. 2011. "Rethinking the Thinking Cap: Ethics of Neural Enhancement Using Noninvasive Brain Stimulation." *Neurology* 76, no. 2: 187–193.

Harman, Gilbert. 1999. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." In *Proceedings of the Aristotelian Society*, 315–331. Aristotelian Society.

Hayes, Steven C., Kelly G. Wilson, Elizabeth V. Gifford, Victoria M. Follette, and Kirk Strosahl. 1996. "Experiential Avoidance and Behavioral Disorders: A Functional Dimensional Approach to Diagnosis and Treatment." *Journal of Consulting and Clinical Psychology* 64, no. 6: 1152.

Hoedlmoser, Kerstin, Thomas Pecherstorfer, Georg Gruber, Peter Anderer, Michael Doppelmayr, Wolfgang Klimesch, and Manuel Schabus. 2008. "Instrumental Conditioning of Human Sensorimotor Rhythm (12–15 Hz) and Its Impact on Sleep as Well as Declarative Learning." *Sleep* 31, no. 10: 1401–1408.

Hsueh, Jen-Jui, Tzu-Shan Chen, Jia-Jin Chen, and Fu-Zen Shaw. 2016. "Neurofeedback Training of EEG alpha Rhythm Enhances Episodic and Working Memory." *Human Brain Mapping* 37, no. 7: 2662–2675.

Hursthouse, Rosalind. 1991. "Virtue Theory and Abortion." *Philosophy & Public Affairs* 20, no. 3: 223–246.

Iyer, Meenakshi B., U. Mattu, J. Grafman, M. Lomarev, S. Sato, and E. M. Wassermann. 2005. "Safety and Cognitive Effect of Frontal DC Brain Polarization in Healthy Individuals." *Neurology* 64, no. 5: 872–875.

Jotterand, Fabrice, and James Giordano. 2015. Real-Time Functional Magnetic Resonance Imaging—Brain-Computer Interfacing in the Assessment and Treatment of Psychopathy: Potential and Challenges. In *Handbook of Neuroethics,* edited by Jens Clausen and Neil Levy, 763–781, Dordrecht: Springer.

Juengst, Eric, and Moseley, Daniel. Spring edition 2016. "Human Enhancement," *The Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta. Available at https://plato.stanford.edu/

Kadosh, Roi Cohen, Neil Levy, Jacinta O'Shea, Nicholas Shea, and Julian Savulescu. 2012. "The Neuroethics of Non-invasive Brain Stimulation." *Current Biology* 22, no. 4: R108–R111.

Kahane, Guy, and Julian Savulescu. 2015. "Normal Human Variation: Refocussing the Enhancement Debate." *Bioethics* 29, no. 2: 133–143.

Karle, James W., Scott Watter, and Judith M. Shedden. 2010. "Task Switching in Video Game Players: Benefits of Selective Attention but Not Resistance to Proactive Interference." *Acta psychologica* 134, no. 1: 70–78.

Kiesel, Andrea, Wilfried Kunde, and Joachim Hoffmann. 2007. "Unconscious Priming According to Multiple SR Rules." *Cognition* 104, no. 1: 89–105.

Klimecki, Olga. M. 2019. The Role of Empathy and Compassion in Conflict Resolution. *Emotion Review*, *11*(4): 310–325.

Kristjánsson, Árni. 2013. "The Case for Causal Influences of Action Videogame Play upon Vision and Attention." *Attention, Perception, & Psychophysics* 75, no. 4: 667–672.

Kristjánsson, Kristján. 2015. *Aristotelian Character Education*. Abington-on-Thames: Routledge.

Li, Renjie, Uri Polat, Fabien Scalzo, and Daphne Bavelier. 2010. "Reducing Backward Masking through Action Game Training." *Journal of Vision* 10, no. 14:33.

Lindenberger, Ulman, Elisabeth Wenger, and Martin Lövdén. 2017. "Towards a Stronger Science of Human Plasticity." *Nature Reviews Neuroscience* 18, no. 5: 261.

Little, Simon, Alex Pogosyan, Spencer Neal, Baltazar Zavala, Ludvic Zrinzo, Marwan Hariz, Thomas Foltynie et al. 2013. "Adaptive Deep Brain Stimulation in Advanced Parkinson Disease." *Annals of Neurology* 74, no. 3: 449–457.

Luchtman, Dirk W., and Cai Song. 2013. "Cognitive Enhancement by Omega-3 Fatty Acids from Childhood to Old Age: Findings from Animal and Clinical Studies." *Neuropharmacology* 64: 550–565.

Lucivero, Federica, and Guglielmo Tamburrini. 2008. "Ethical Monitoring of Brain-Machine Interfaces." *AI & Society* 22, no. 3: 449–460.

Mak, Joseph N., and Jonathan R. Wolpaw. 2009. "Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects." *IEEE Reviews in Biomedical Engineering* 2: 187–199.

Miranda, Robbin A., William D. Casebeer, Amy M. Hein, Jack W. Judy, Eric P. Krotkov, Tracy L. Laabs, Justin E. Manzo et al. 2015. "DARPA-Funded Efforts in the Development of Novel Brain–Computer Interface Technologies." *Journal of Neuroscience Methods* 244: 52–67.

Montmarquet, James A. 1987. "Epistemic Virtue." *Mind* 96, no. 384: 482–497.

Nan, Wenya, João Pedro Rodrigues, Jiali Ma, Xiaoting Qu, Feng Wan, Pui-In Mak, Peng Un Mak, Mang I. Vai, and Agostinho Rosa. 2012. "Individual Alpha Neurofeedback Training Effect on Short Term Memory." *International Journal of Psychophysiology* 86, no. 1: 83–87.

Niemiec, Christopher P., and Richard M. Ryan. 2009. "Autonomy, Competence, and Relatedness in the Classroom: Applying Self-Determination Theory to Educational Practice." *Theory and Research in Education* 7, no. 2: 133–144.

Niv, Sharon. 2013. "Clinical efficacy and potential mechanisms of neurofeedback." *Personality and Individual Differences* 54, no. 6: 676–686.

Nussbaum, Martha C. 2001. *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. Cambridge: Cambridge University Press.

O'Brolcháin, Fiachra, and Bert Gordijn. 2015. "Ethics of Brain–Computer Interfaces for Enhancement Purposes." *Handbook of Neuroethics,* edited by Jens Clausen and Neil Levy, 1207–1226, Springer, Dordrecht.

Pascual-Leone, Alvaro, Catarina Freitas, Lindsay Oberman, Jared C. Horvath, Mark Halko, Mark Eldaief, Shahid Bashir et al. 2011. "Characterizing Brain Cortical Plasticity and Network Dynamics across the Age-Span in Health and Disease with TMS-EEG and TMS-fMRI." *Brain topography* 24, no. 3–4: 302.

Patsenko, Elena G., et al. 2019. "Mindfulness Video Game Improves Connectivity of the Fronto-Parietal Attentional Network in Adolescents: A Multi-modal Imaging Study." *Scientific Reports* 9, no.1: 1–8.

Persson, Ingmar, and Julian Savulescu. 2012. *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Pohl, Carsten, Wilfried Kunde, Thomas Ganz, Annette Conzelmann, Paul Pauli, and Andrea Kiesel. 2014. "Gaming to See: Action Video Gaming Is Associated with Enhanced Processing of Masked Stimuli." *Frontiers in Psychology* 5: 70.

Qin, Shaozheng, Erno J. Hermans, Hein J. F. van Marle, Jing Luo, and Guillén Fernández. 2009. "Acute Psychological Stress Reduces Working Memory-Related Activity in the Dorsolateral Prefrontal Cortex." *Biological Psychiatry* 66, no. 1: 25–32.

Reiner, Miriam, Roman Rozengurt, and Anat Barnea. 2014. "Better than Sleep: Theta Neurofeedback Training Accelerates Memory Consolidation." *Biological Psychology* 95: 45–53.

Richter, Thalia, Alexander J. Shackman, Tatjana Aue, and Hadas Okon-Singer. 2019. "The Neurobiology of Emotion–Cognition Interactions." In *Cognitive Dimensions of Major Depressive Disorder,* edited by Bernhard T. Baune and Catherine Harmer, 171–183. Oxford and New York: Oxford University Press.

Roberts, Robert C., and W. Jay Wood. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford University Press on Demand.

Roque, Nelson A., and Walter R. Boot. 2018. "Action Video Games DO NOT Promote Visual Attention." In *Video Game Influences on Aggression, Cognition, and Attention*, 105–118. Cham: Springer.

Rozengurt, Roman, Anat Barnea, Sunao Uchida, and Daniel A. Levy. 2016. "Theta EEG Neurofeedback Benefits Early Consolidation of Motor Sequence Learning." *Psychophysiology* 53, no. 7: 965–973.

Sandel, Michael J. 2009. *The Case against Perfection*. Cambridge, MA: Harvard University Press.

Schermer, Martije. 2008. On the Argument that Enhancement Is "Cheating." *Journal of Medical Ethics*, *34*(2), 85–88.

Schwitzgebel, Eric, and Joshua Rust. 2009. "The Moral Behaviour of Ethicists: Peer Opinion." *Mind* 118, no. 472: 1043–1059.

Selart, Marcus, and Svein Tvedt Johansen. 2011. "Ethical Decision Making in Organizations: The Role of Leadership Stress." *Journal of Business Ethics* 99, no. 2: 129–143.

Sellaro, Roberta, Belle Derks, Michael A. Nitsche, Bernhard Hommel, Wery P. M. van den Wildenberg, Kristina van Dam, and Lorenza S. Colzato. 2015. "Reducing Prejudice through Brain Stimulation." *Brain Stimulation* 8, no. 5: 891–897.

Shickle, Darren. 2000. "Are 'Genetic Enhancements' Really Enhancements?" *Cambridge Quarterly of Healthcare Ethics* 9, no. 3: 342–352.

Shih, Jerry J., Dean J. Krusienski, and Jonathan R. Wolpaw. 2012. "Brain-Computer Interfaces in Medicine." In *Mayo Clinic Proceedings*, 87, no. 3: 268–279. Elsevier.

Simons, Daniel J., Walter R. Boot, Neil Charness, Susan E. Gathercole, Christopher F. Chabris, David Z. Hambrick, and Elizabeth A. L. Stine-Morrow. 2016. "Do "Brain-Training" Programs Work?" *Psychological Science in the Public Interest* 17, no. 3: 103–186.

Solon, Olivia. "Elon Musk Says Humans Must Become Cyborgs to Stay Relevant. Is He Right?" *The Guardian*, Feb 15, 2017 https://www.theguardian.com/technology/2017/feb/15/elon-musk-cyborgs-robots-artificial-intelligence-is-he-right

Song, Dong, Brian S. Robinson, Robert E. Hampson, Vasilis Z. Marmarelis, Sam A. Deadwyler, and Theodore W. Berger. 2016. "Sparse Large-Scale Nonlinear Dynamical Modeling of Human Hippocampus for Memory Prostheses." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, no. 2: 272–280.

Sorabji, Richard. 1980. *Necessity, Cause, and Blame: Perspectives on Aristotle's Theory*. Chicago, IL: University of Chicago Press.

Stahnisch, Frank W., and Robert Nitsch. 2002. "Santiago Ramón y Cajal's Concept of Neuronal Plasticity: The Ambiguity Lives On." *Trends in Neurosciences* 25, no. 11: 589–591.

Stojanoski, Bobby, Kathleen M. Lyons, Alexandra A. A. Pearce, and Adrian M. Owen. 2018. "Targeted Training: Converging Evidence against the Transferable Benefits of Online Brain Training on Cognitive Function." *Neuropsychologia* 117: 541–550.

Strobach, Tilo, Peter A. Frensch, and Torsten Schubert. 2012. "Video game practice optimizes executive control skills in dual-task and task switching situations." *Acta psychologica* 140, no. 1: 13–24.

Swanton, Christine. 2003. *"Virtue Ethics: A Pluralistic View."* Oxford/New York: Oxford University Press.

Torous, John, Patrick Staples, Elizabeth Fenstermacher, Jason Dean, and Matcheri Keshavan. 2016. "Barriers, Benefits, and Beliefs of Brain Training Smartphone Apps: An Internet Survey of Younger US Consumers." *Frontiers in Human Neuroscience* 10: 180.

Tversky, Amos, and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5, no. 2: 207–232.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211, no. 4481: 453–458.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.

Van Gerven, Marcel, Jason Farquhar, Rebecca Schaefer, Rutger Vlek, Jeroen Geuze, Anton Nijholt, Nick Ramsey et al. 2009. "The Brain–Computer Interface Cycle." *Journal of Neural Engineering* 6, no. 4: 041001.

van Ravenzwaaij, Don, Wouter Boekel, Birte U. Forstmann, Roger Ratcliff, and Eric-Jan Wagenmakers. 2014. "Action Video Games Do Not Improve the Speed of Information Processing In Simple Perceptual Tasks." *Journal of Experimental Psychology: General* 143, no. 5: 1794.

Vernon, David, Tobias Egner, Nick Cooper, Theresa Compton, Claire Neilands, Amna Sheri, and John Gruzelier. 2003. "The Effect of Training Distinct Neurofeedback Protocols on Aspects of Cognitive Performance." *International Journal of Psychophysiology* 47, no. 1: 75–85.

Vlek, Rutger J., David Steines, Dyana Szibbo, Andrea Kübler, Mary-Jane Schneider, Pim Haselager, and Femke Nijboer. 2012. "Ethical Issues in Brain–Computer Interface Research, Development, and Dissemination." *Journal of Neurologic Physical Therapy* 36, no. 2: 94–99.

Wang, Jinn-Rong, and Shulan Hsieh. 2013. "Neurofeedback Training Improves Attention and Working Memory Performance." *Clinical Neurophysiology* 124, no. 12: 2406–2420.

Watanabe, Yoshifumi, Elizabeth Gould, and Bruce S. McEwen. 1992. "Stress Induces Atrophy of Apical Dendrites of Hippocampal CA3 Pyramidal Neurons." *Brain Research* 588, no. 2: 341–345.

Wegner, Daniel M., David J. Schneider, Samuel R. Carter, and Teri L. White. 1987. "Paradoxical Effects of Thought Suppression." *Journal of Personality and Social Psychology* 53, no. 1: 5.

Wu, Sijing, and Ian Spence. 2013. "Playing Shooter and Driving Videogames Improves Top-Down Guidance in Visual Search." *Attention, Perception, & Psychophysics* 75, no. 4: 673–686.

Yuste, Rafael, et al. 2017. "Four Ethical Priorities for Neurotechnologies and AI." *Nature News* 551, no. 7679: 159.

Zilles, Karl. 1992. "Neuronal Plasticity as an Adaptive Property of the Central Nervous System." *Annals of Anatomy-Anatomischer Anzeiger* 174, no. 5: 383–391.

CHAPTER 29

TOWARD AN EXISTENTIAL AND EMANCIPATORY ETHIC OF TECHNOLOGY

CHARLES ESS

> For what the highest degree may be at which [hu]mankind may have to come to a stand, and how great a gulf may still have to be left between the idea and its realization, are questions which no one can, or ought to, answer. For the issue depends on freedom; and it is in the power of freedom to pass beyond any and every specified limit [*Grenze*].
>
> Kant [1781, 1787] 1965, 312; A317 B374.[1]

1. Introduction

I offer a broad outline of what good lives of human meaning and flourishing can look like in an era defined by our technologies and, specifically, media technologies. The essay proceeds in four parts. The first begins developing a *philosophical anthropology*, an account of *being human*. The ancient *Epic of Gilgamesh* introduces the central existential themes of confronting our mortality and the challenge of discerning and creating meaningful lives as embodied beings—and as *relational autonomies* that cohere with care ethics, virtue ethics, and deontological emphases on respect and equality. Embodiment further introduces a thematic focus here on *dualisms*—mind/body, male/female, (human) nature/technology—and the importance of overcoming these. I argue that we may learn from *Gilgamesh* and its wisdom sayings: but to apply these in our contemporary era, as far more defined by our technologies, requires a turn in the second section to the philosophical and theological origins of modern technology. I foreground here *emancipation* as the defining aim of the Enlightenment and modern technology—first of all, for the sake of the deontological goal of becoming a self-governing freedom (*autonomy*).

To be sure, such Enlightenment conceptions of autonomy and emancipation are deeply fraught, as feminist and postcolonial criticisms (among others) have laid bare. As Nikita Dhawan succinctly puts it, "Along with progress and emancipation, [Enlightenment Reason] has brought colonialism, slavery, genocide, and crimes against humanity" (Dhawan 2014, 9). I will briefly discuss these critiques and their suggestions of how we can sustain and enhance these central notions in a "decolonized Enlightenment" that retains its originary aim of overcoming "equalities and injustice in a postcolonial world" (2014, 10). Broadly, the turns argued here from dualism to non-dualism—specifically the feminist notion of *relational autonomy* along with foregrounding Romanticism and Enlightenment entanglements—thereby incorporate and resonate with these revisions of Enlightenment, specifically Kantian understandings of freedom and emancipation.

In these directions, I then argue that a modern interest in emancipation can be understood in two distinct ways. Briefly, what we can call the Baconian-Cartesian program for a modern technology aims at nothing less than eliminating labor and overcoming death: this program, however, rests on yet another set of dualisms (mind/body, human/nature)—ones that issue in fears of "autonomous technology" and Frankenstein monsters, including persistent figures of "techno-femme fatale" (Consalvo 2004) robots that inevitably turn on and destroy us. By contrast, I argue, following Mark Coeckelbergh, that we can overcome these dangers by moving beyond such dualisms, including the correlative dualism of Enlightenment Reason versus Romantic emphases on emotion and the transcendent. The resulting *complementarity* of Reason and Romanticism thereby crucially expands our anthropology and ethical toolkit.

Third, I expand on the existential themes of the *Gilgamesh* by way of contemporary existentialism and virtue ethics. As our lives are inextricably interwoven with technologies, especially media technologies, we examine how existential media studies (Lagerkvist 2018) foregrounds Karl Jaspers' central existential concepts of *Existenz* and the *limit-situation* (*Grenzsituation*) as fruitfully applicable to existentialism in a digital era. Shannon Vallor's (2016) "technomoral virtues," including the key virtue of courage, then come into play here as a final ethical component, one also grounded in existential thought.

Fourth, I return to the ancient stories of Gilgamesh and the woman in the Garden: courage is the virtue needed for confronting our own mortality and thereby taking on primary responsibility for discerning and creating meaning in our lives—as it is also the virtue required for moving beyond child-like obedience to authorities (through resistance and disobedience), in order to acquire the Kantian/relational autonomy central to emancipation and thereby the possibilities of cultivating our own ethical and moral judgments (*phronēsis*) in good lives of flourishing. I close by sketching some specific ways in which we might do so—ways that, I hope, instantiate Romantic-Enlightenment conceptions of "romantic machines" that genuinely serve our individual and collective emancipation and flourishing.

2. A Philosophical Anthropology

I take up a *philosophical anthropology* that responds to the primary question: who/what are we as human beings? I start with the ancient *Epic of Gilgamesh*.[2] *Gilgamesh* sets the existential stage for understanding our being human: our "growing up," our learning to become human requires our coming to grips with our mortality, the reality of our own death. *Gilgamesh* thereby grounds central notions of *embodiment* and *relational autonomy*, conjoined with the primary ethical frameworks of *virtue ethics, care ethics*, and *deontological ethics*.

The ancient *Epic of Gilgamesh* recounts how the death of his boon companion Enkidu forces the seemingly invincible warrior-king Gilgamesh to confront his own mortality. Gilgamesh ultimately fails in his extraordinary efforts to achieve immortality. But this failure is countered by an unexpected gift of wisdom from a young woman (or goddess), the wine-maker Siduri:

> Gilgamesh, where are you hurrying to? You will never find that life for which you are looking. When the gods created man they allotted to him death, but life they retained in their own keeping. As for you, Gilgamesh, fill your belly with good things; day and night, night and day, dance and be merry, feast and rejoice. Let your clothes be fresh, bathe yourself in water, cherish the little child that holds your hand, and make your wife happy in your embrace: for this too is the lot of man.
>
> (Sandars 1972, 102)

And so Gilgamesh learns to grow up. Specifically, as Siduri admonishes in her final lines, Gilgamesh becomes more caring for those whom, as weaker than himself, he once exploited without limit.

For all the differences between our age and ancient Sumeria, we share with Gilgamesh the deeply human tendency to deny our mortality. But central to our "growing up," becoming more caring and wiser human beings, is precisely the necessity of moving beyond such denial to recognition and acceptance—and thereby finding and creating *meaning* through care for others as well as for ourselves. Gilgamesh is thus a primordial *existential* text—where existentialism enjoins precisely our confronting a mortality we'd much rather deny.

We will see that countering our otherwise all-too-prevailing tendencies and efforts to deny our mortality requires *courage* first of all—a virtue or capability that must be practiced. To make one's spouse or partner happy in one's embrace also requires *care*—care for the Other as a full and complete human being in her or his own right, not merely "meat" useful only for the satisfaction of one's own sexual desires. Care is similarly required for "cherish[ing] the little child that holds your hand"—again, responding to someone far more vulnerable and in need of care with care. *Deontological* ethics begins with the insistence on human beings as rational autonomies or freedoms who thereby deserve to always be treated "as ends, never as means only" (Kant [1785] 1993, 43): as we have come to say, as rights-holders to be treated with respect as equals.

That Gilgamesh comes to these ethical practices and sensibilities is a central lesson in the *Epic*. At the beginning of the story, as more powerful than any other human being, Gilgamesh is possessed of unbounded arrogance: "No son is left with his father, for Gilgamesh takes them all, even the children . . . His lust leaves no virgin to her lover, neither the warrior's daughter nor the wife of the noble" (Sandars 1972, 62). His final acknowledgment of his own mortality, catalyzed by the death of his friend and comrade Enkidu, is critical to his coming to accept Siduri's wisdom saying and to begin to exercise in practice the virtue of courage, care, and respect for the Others he once only exploited without limit. Thus it is arguable that Siduri is enjoining Gilgamesh towards a virtue ethics and a care ethics—ones that entail a deontological respect for others as well.

2.1 Embodiment

To state the obvious: Gilgamesh encounters the realities of mortality only because he is an *embodied* human being. The primal fact of embodiment is that bodies sicken, age, and die. But just as our embodiment immediately entails vulnerability and death—so it entails multiple possible pleasures. The hedonistic plunge into pleasure as a compensation for death is, of course, commonplace. And certainly, the first elements of Siduri's advice to Gilgamesh stress *bodily* pleasures: "As for you, Gilgamesh, fill your belly with good things; day and night, night and day, dance and be merry, feast and rejoice. Let your clothes be fresh, bathe yourself in water." But these pleasures are immediately conjoined with specific *ethical* injunctions of care and respect: "cherish the little child that holds your hand, and make your wife happy in your embrace: for this too is the lot of man" (Sandars 1972, 102).

Taking embodiment first: obviously, we "have" bodies. But to say it this way—in the way we might say "I have a pen"—implies that there is an agent, an "I" or a self, who somehow stands apart from the body. Insofar as we speak and think this way—we do so as inheritors of a stubborn Christian and then Cartesian *dualism* that sunders the soul or rational self from body (Ess 2017, 87). In contrast, as especially phenomenologists in the twentieth century began to make clear, we *are* our bodies in central ways. So Maurice Natanson writes:

> *I am my body*. There is no distance between my hand and its grasping. [. . .] Instead of the common-sense way of thinking of the body in space at some time, I am a corporeality Here and Now whose being in the world is disclosed to me as mine.
>
> (1970: 11, italics in original)

This is not to reduce human being to body only. On the contrary, phenomenologists also foreground a *first-person phenomenal consciousness*, which includes (conscious) attention to objects, whether in an external world and/or within consciousness itself. Moreover, we are aware of having at least some choice and control over such attention—where such control and choice are central components of an underlying freedom or autonomy. Embodiment here means an understanding of self and body as oftentimes phenomenally different, oftentimes phenomenally indistinguishable—but finally a *non-dualistic* insistence on body as inextricable from our identity, being, and knowledge (cf. Ruddick 1975).

Such embodiment, nonetheless, immediately entails the existential point: our bodies are mortal, as well as vulnerable in other ways. We, like Gilgamesh, are likely to do all we can to deny our mortality. One strategy for doing so is to posit something like a Gnostic or Christian soul—or, later, a Cartesian rational mind—that is radically divorced from our bodies and so can live on after the body dies. Embodiment denies such dualisms—and thereby short-circuits such ontologies as possible escapes from mortality.

Moreover, embodiment is central to how we *know*. Again, *contra* a Cartesian dualism that would split conscious mind from unconscious body—theories of "the extended mind," embodied mind, and their supporting "philosophy of embodiment" (Boden 2006, 1404–1407) foreground the multiple ways in which we are aware of and know the world both through our bodies and, indeed, our technologies, for example, the multiple objects we create and use with our bodies as part of our cognitive processes (e.g., counting knots in strings, writing with pens and paper, and/or interacting with computational devices: cf. Wilson and Foglia 2017).

2.2 Relational Selves and Ethics

Again, Siduri enjoins Gilgamesh to "cherish the little child that holds your hand, and make your wife happy in your embrace: for this too is the lot of man" (Sandars 1972, 102): as we have seen, these injunctions point to a deontological ethics and an ethics of care. I further read these as an affirmation of our deep *relationality* and thereby *relational* selves.

By contrast, Christian and Cartesian claims regarding soul or mind as separate from body tend to further portray the self as a psychic atom—that is, a singular and distinctive *individual* self, one defined first of all by its *difference* from everything around it, including other human beings. Certainly, such conceptions are vitally important as they become more developed especially in Kant as *rational autonomies*, beings that are free to determine their own ends and laws via reason: such autonomies ground modern theories of democratic rights, processes, and norms such as fundamental respect and equality (e.g., Ess 2010). Taken to extremes, however, these atomistic conceptions become highly problematic, beginning in the political sphere. For Thomas Hobbes (1588–1679), these psychic atoms are centrally desire-driven and self-interested: hence our relationships with one another in a "state of nature" can only be competition and conflict—a "warre of every man against every man" in short (Hobbes 1651, 57). The only counter to such chaos is the absolute monarch, the authoritarian leader to whom everyone else cedes their basic rights in a more orderly and thereby more commodious society (e.g., Baier 1987). Such assumptions regarding human beings and human nature thus undergird authoritarian politics, norms and virtues, in direct conflict with democratic ones.

More recent research and reflection—in philosophy, the social sciences, and neuroscience—rather emphasize a *relational self*. As a prime example: contemporary media studies focusing on social networking sites (SNSs) such as Facebook and social media such as Twitter, Instagram, Snapchat, and so on, consistently invoke the sociologist Erving Goffman who stressed how multiple relationships define our identity and so require our on-going curation of the diverse presentations of the selves that fit specific relationships (1959; e.g., Lomborg 2012; Ess and Fossheim 2013).

Such relational selfhood also threatens, however, autonomy and democracy as it utterly depends upon acknowledgement and affirmation from the multiple others whose relationships define it. Indeed, relational selves emerge in conjunction with hierarchical and authoritarian societies; the very existence of such selves depends on *obedience* to one's many superiors and their *authority* (e.g., MacIntyre 1994, 190; Ess and Fossheim 2013, 48f.). Our shifting towards more relational selves, especially via social and related media, thereby runs the risk of elevating obedience and authoritarianism at the cost of democratic freedoms and rights as resting on more individual conceptions of selfhood. It may not be accidental that the rise of social media and relational selves is accompanied by growing far-right, more authoritarian politics in the West.

To avoid the risks of both extreme individualism and relationality, recent feminist theorists have developed conceptions of *relational autonomy*: these conjoin the multiple realities and benefits of relationality—especially as enunciated in feminist ethics and *ethics of care* (Tappolet 2014; Westlund 2014)—with autonomy as core to *deontological ethics* and thereby democratic norms and processes. Moreover, Andrea Veltman and Mark Piper invoke the central aims of *virtue ethics*, asserting that "autonomy is one primary good among others that a person needs to live a good life or to achieve human flourishing" (2014, 2). Specifically, Veltman (2014) conjoins virtue ethics with a Kantian deontological account of autonomy that grounds respect for persons as a primary value. Such a conjunction thereby reiterates and expands upon Kant's own development of deontology and conjunction with his virtue ethics (Kant [1797] 1964).[3] Moreover, as we will explore more fully in the next section, incorporating these notions of relational autonomy is a central way of taking on board here both feminist and postcolonial critiques of Enlightenment conceptions of reason and emancipation.

So Gilgamesh learned to move beyond his initial warrior self, for whom others could only be enemies to be defeated through violence or objects of sexual exploitation. Learning to care for—"cherish"—both the little child and the interests and happiness of his partner is a primary *virtue*, one that emphasizes the role of our relationships with one another in pursuing good lives of flourishing. To be sure, Kantian deontology, care ethics, and notions of relational autonomy are considerably more sophisticated contemporary developments: but they clearly resonate with Siduri's wisdom saying and suggest that for all the differences separating us from the ancient Sumerian warrior, we nonetheless share the concern with living good lives of flourishing in the face of our mortality.

One of our differences from Gilgamesh is the role and significance of contemporary technologies—specifically, as these come to define both *who we are and may become as human beings* as well as *the conditions of our existence*. To see how this is so—first of all, for the sake of bringing to the foreground the modern thematic of *emancipation* via new science and technologies as central to the pursuit of good lives as embodied beings—we first must recall the philosophical as well as theological backgrounds to the emergence of modern technology.

3. Historical and Theological Backgrounds of Modern Technology

3.1 Dualism, Original Sin and the Aims of Modern Technology

This means first that we must move beyond yet one more common dualism—namely, the assumption that our technologies are somehow radically divorced from us as their creators and users. This assumption is driven in modernity in large measure by the Romantic reaction to the new science and (excessive) rationalism of the Enlightenment. With antecedents in German Romanticism (Coeckelbergh 2017a, 27–41), English Romanticism and specifically Mary Shelley's novel *Frankenstein: or, a Modern Prometheus* ([1818] 2003) establish the now classic themes defining our fears of technology: out of *hubris* or excessive pride in our new sciences and possibilities, we overstep our human bounds and take on the god-like role of creating new life. But this new being—"thy Adam" (Shelley [1818] 2003, 61; cf. 81f.)—proves frightening to us, we reject it, and it then turns on us in murderous revenge. From here, there is a direct line to what Langdon Winner characterized as "autonomous technology" and technology out of control as a central theme in twentieth century philosophy of technology and popular culture (1977).

The reference to Adam indexes how Shelley's novel implicates the religious and theological backgrounds of modern technology (Ess 2017). Most briefly, the primary goal of modern technology, as articulated by Francis Bacon (1561–1626) and René Descartes (1596–1650), is to exploit the new sciences in order to make human beings "masters and possessors of nature" (Descartes [1637] 1972, 119; Ess 2017, 87). But what to do with our new-found powers? Descartes identifies two primary aims: the elimination of labor and the pains of old age—perhaps even death ([1637] 1972, 119f.; Ess 2017, 87).

These aims are taken for granted by moderns. But for his contemporaries—people still steeped in the worldview and Scriptures of Western Christianities—Descartes' aims immediately recalled the Garden of Eden story (Gen 2.4–3.24), as interpreted through St. Augustine (354–430). Augustine's reading reversed the earlier Jewish and Christian understandings of the story as—like Gilgamesh—a story of how human beings grow up. Specifically, on these earlier readings, the woman[4] must *disobey* a parent-like authority in order to acquire autonomy and ethical judgment, the "knowledge of good and evil" (Gen. 2.17; 3.5; 3.22). The woman thereby leads the way to human maturity—maturity that requires the *virtue* of *disobedience*. Augustine turns this reading upside down, portraying the woman's disobedience to male authority as the primal—Original—sin and thereby rendering her as primarily responsible for its well-known consequences: expulsion from the Garden into a life marked by labor, death, and pain in childbirth (Gen. 3.16–19). *Contra* the existential and embodied joys of good food and drink, and enjoying "life with the wife whom you love" (Ecclesiastes 9.9)—(Augustinian) "salvation" from this human condition is instead to be sought for the immortal soul as radically separate from the mortal body. Given this Augustinian reading, the Cartesian aims for modern science and technology are clearly *hubris* and blasphemy: modern technology, so defined, aims at nothing less than overturning the consequences of human disobedience by making us gods indeed who no longer labor, sicken, or die (Ess 2017, 87f.).

This religious and theological background remains strikingly trenchant in our ostensibly more secular age: so the Frankenstein story inspires twentieth and twenty-first century explorations of robots. From Fritz Lang's *Metropolis* (1927) to Eva (a conflation of Eve and Adam) in *Ex Machina* (2015), robots consistently take on seductive female form—and thereby the disobedient and destructive persona of the woman in Augustine's reading. Surprisingly, this "techno-femme fatale" (Consalvo 2004) hence centers our otherwise secular understandings of contemporary technologies on a strikingly religious image—thereby conjuring an interpretation of the Garden story that has justified male domination and violence against women for millennia (cf. Ess 1995). For my part: any project of pursuing good lives of flourishing rests upon fundamental deontological values of respect and equality, including gender equality. Bluntly: there is no place in such a project for the on-going presence and influence of Augustine's demonization of women, sexuality, body—and thereby nature.

Specifically: a philosophical anthropology that endorses non-dualistic conceptions of selfhood and embodiment (recall Natanson 1970, 11) immediately entails rejecting an Augustinian soul-body dualism, as well as its Cartesian version, that is, the mind-body split. This is likewise to move beyond an Augustinian *contemptus mundi* ("contempt for the world," including contempt for women) that roots the Cartesian human mastery and possession of nature (Ess 2004). Replacing these dualisms with more non-dualistic conceptions is further coherent with broad contemporary shifts towards ecological philosophies that emphasize human-nature entanglement. This further implies rejecting related dualisms, such as those assumptions of men as rational and women as emotive (hence their alleged irrational disobedience of rational authority, etc. Ess 2017, 87)—and thereby to move away from the traditional patriarchies that they undergird. This coheres directly, then, with the deontological insistence on respect and equality due to all persons as (relational) autonomies. Last but not least, these non-dualistic conceptions directly undercut the Us versus Them, Master/Slave, and related dichotomies used to justify slavery, colonization, imperialism and so on—that is, part and parcel with feminist and postcolonial critiques of the worst outcomes of Enlightenment reason. Indeed, we are about to see how the entanglements between early Enlightenment and *Romanticism* moves precisely in these feminist and postcolonial directions at the outset.

3.2 Enlightenment, Romanticism, and New Romantic Cyborgs

Mark Coeckelbergh (2017a) further extends this philosophical anthropology and its correlative ethical dimensions through his historical exploration of the Enlightenment (including Kant) and Romanticism. This account corrects a prevailing assumption of a sharp opposition between Enlightenment rationalism vis-à-vis[5] Romanticism and thereby moves us further beyond the (male) reason versus (female) affect dualism. This non-dualistic understanding rather insists that our being/becoming human is inextricable with our technologies.

First, Coeckelbergh challenges a prevailing assumption that places Romanticism and the Enlightenment in opposition. Certainly, early Romanticism reacted strongly against "Enlightenment rationalism, scientific objectivism, disenchantment, and attempts to crush religion" by attempting "to revive and liberate subjective feeling and emotion, passion, horror, and melancholy" (Coeckelbergh 2017a, 1). A closer look at their early history shows, however, that "the line between Enlightenment and Romanticism was thinner than one might expect—even for philosophers like Kant, who was haunted by both rationalism and mysticism" (2017a, 13). Specifically, Kant establishes a sharp distinction between the *phenomenal* world, that is, the world as it is known to us via our senses and our cognitive processes, vis-à-vis its *noumenal* underpinnings which thereby remain forever beyond human knowledge in a strict sense: this distinction helps root the Romantic "emphasis on the unknowable—there are limits to knowledge" (Coeckelbergh 2017a, 33). As Kant himself stated in the second edition of his *Critique of Pure Reason* with such concepts as the *noumena/phenomena* distinction he sought to establish the *limits* of reason in order to make room for faith (Kant [1781, 1787] 1965, 29; B xxx). For Kant, such faith must remain famously "within the bounds of reason alone" (Kant [1793] 1968): but as Romantics and contemporary existentialists recognize, setting limits to reason and its ways of knowing renders "transcendence"—minimally, that which is beyond our ordinary modes of knowing—possible. Specifically, this Kantian background becomes central to two of the primary *existential* conceptions taken up in the next section—namely, Karl Jaspers' understandings of the "limit-situation" (*Grenzsituation*) and *transzendenz* ("transcendence").

Coeckelbergh further documents a Romantic interest in machines and technology, in part as Romantics saw in these, echoing Descartes, possibilities for *liberation*. The early Marx, for example, highlights themes of "life, freedom, and spontaneity" as central grounds for critiquing how labor as organized under industrial capitalism alienates us from our freedom and human being (Coeckelbergh 2017a, 36). In the later "material romanticism" of Marx and Engels in *The German Ideology* (1846), these authors observe that "slavery cannot be abolished without the steam-engine" ([1846] 1976, 38, in Coeckelbergh 2017a, 37). Coeckelbergh points to still further Romantics who take up technology "as a tool for social change" (2017a, 37) and a political focus emphasizing democracy, a "tolerance and pluralism" regarding all cultures, and in some cases, socialism" (2017a, 55).

A Romantic-Enlightenment vision thus emerges here of democracy, pluralism, and emancipation, in part by way of technologies that reduce or eliminate enforced labor—for women, enslaved peoples, and nature itself as understood as something far greater than mere "matter" as allegedly separate from and thus to be mastered by human rationality in the Cartesian agenda. As critical theorists, feminists, and postcolonial critiques have made clear, however, some versions of Enlightenment reason and emancipation are deeply fraught. Theodore Adorno and Horkheimer's *Dialectic of Enlightenment* ([1947] 2002), as a start, confronts the ashes and atrocities of twentieth century fascism and genocide as a "disaster triumphant" of an allegedly enlightened Earth (Zuidervaart 2015, 4). Starting with critiques of Bacon and Descartes' conceptions of a reason that becomes instrumentalized and thereby irrational, Horkheimer and Adorno seek to move beyond a triple domination: "the domination of nature by human beings, the domination of nature within human beings, and, in both of these forms of domination, the domination of some human beings by others" (Zuidervaart 2015, 4). Similarly, postcolonial scholars such as Nikita Dhawan succinctly observe that "Along with progress and emancipation, [Enlightenment] reason] has brought colonialism, slavery, genocide, and crimes against humanity" (2014, 9).

As we have seen, however, the feminist conception of relational autonomy, along with allied non-dualisms, starting with reason-affect and extending through human-nature, take on board at least some of these critiques. By the same token, Dhawan characterizes the aim of at least some postcolonial scholars to show the coercive, repressive, and exploitive sides of "discourses of transnational justice, human rights, and democracy" (2014, 10)—while retaining, if necessarily transforming, the central impulses towards emancipation and equality (2014, 10) As Susan Buck-Morss put it early on:

> If the historical facts about freedom can be ripped out of the narratives told by the victors and salvaged for our own time, then the project of universal freedom does not need to be discarded but, rather, redeemed and reconstituted on a different basis. (2000, 865)

Here we can see that these feminist and postcolonial insights, however much they are rooted in the experiences and voices of "Others" who have been excluded and exploited in nineteenth and twentieth century manifestations of an instrumental reason, at least cohere with especially Romantic emphases at the outset on emancipation not merely for rational (white) males—but precisely for women, enslaved peoples, and Nature itself.

More broadly: in terms of a philosophical anthropology, Coeckelbergh's account of a contemporary complementarity between reason and affect, Enlightenment and Romanticism, issues in his "new Romantic cyborgs," that is, human beings inextricably interwoven with our technologies (2017, 15). At least, insofar as these technologies emerge as "romantic machines," machines of aesthetic pleasure aimed at emancipation of human beings, they thereby serve to preserve and enhance our (relational) autonomy and a non-dualistic understanding the human-technology relationship. Preserving—perhaps enhancing—the (relational) autonomy of such Romantic cyborgs is to sustain a primary condition for good lives of flourishing (Veltman and Piper 2014, 2).[6]

These new Romantic cyborgs, in short, cohere with a deontological and virtue ethics. This makes sense especially as they would seem to be human beings who can be characterized as embodied and relational, as marked by an embrace of and interconnection with nature and technology, in contrast with the Cartesian subject seeking to master and possess an ostensibly inferior natural order via technology. Insofar as this is so, these new Romantic cyborgs echo and extend the philosophical anthropology begun here with Gilgamesh. How far, however, are they thereby likely to be *existential* Romantic cyborgs?

4. Good Lives and Emancipation in a Technological Era

4.1 Existentialism and Ethics

Gilgamesh introduces for us primary existential themes: the confrontation with our death as embodied beings as potentially leading to our discerning and creating meaning through relationships of care and deontological respect. Modern existentialism, while certainly rooted in older traditions (including, as we have seen, the Wisdom sayings preserved in the Hebrew Bible), begins in the early nineteenth century with Søren Kierkegaard (1813–1855), extends through Friedrich Nietzsche (1844–1900), and then unfolds more fully in the 1940s–1970s (e.g., Solomon 1972). To be sure, numerous philosophers usually counted with existential movements—first of all, Martin Heidegger—grappled in important ways with technology. Contemporary existentialism, emerging primarily in the Scandinavian and Northern European context, is distinctive, however: it begins precisely within considerations of what it means to be human in a technological age—specifically, an era in which our being human is inextricably interwoven with and shaped by our media and communication technologies.

The foremost exponent of a contemporary existentialism, Amanda Lagerkvist, points out specifically that digital media are "irreducibly *existential media*" (2018, 1). Stated differently, the internet, with all of its venues and affordances as increasingly definitive of our lives, is thereby the "existential terrain *par excellence* [as it] provides avenues for exploring the fundamental human condition of being faced with the contingency, absurdity and simultaneous quest for profundity and meaning in our lives" (Lagerkvist 2018, xi). Lagerkvist grounds her analyses in part on the foundational work of media scholars such as John Durham Peters (2015). This re-emergence and transformation of twentieth century existentialism has also been catalyzed by contemporary "Death Online" research and scholarship. In sharp contrast with earlier tendencies in digital venues to deny the reality of human mortality (Ess 2011), what is sometimes called "Grief 2.0" (Hovde 2016) includes attention to new practices of "digitally mediated grieving and memorialising," digital "afterlife" and so on (DORS4 2018).

Lagerkvist takes up especially the insights of Kierkegaard and Karl Jaspers, along with Heidegger, Levinas, and de Beauvoir, among others (Lagerkvist 2018). Specifically, Lagerkvist and Yvonne Andersson foreground Jaspers' central concept of *Existenz*. Lagerkvist and Andersson characterize *Existenz* as emphasizing "the frailty of human existence": such frailty is manifested in "experiences of loss, ill health, and suffering" (Lagerkvist and Andersson 2017, 554). At the same time, however, Jaspers further emphasizes how these experiences may also serve as "sources of fecundity" (Jaspers [1932] 1970, cited in Lagerkvist and Andersson 2017, 554). Specifically, this fecundity is central to Jaspers' conception of the *limit-situation* (*Grenzsituation*). Lagerkvist and Andersson observe that:

> For Jaspers, the limit-situation—of loss, death, crisis, guilt, conflict, and love—is vital since it requires of the individual human being to act and entails the possibility of realizing one's "Existenz": "the limit-situation of being definite calls upon Existenz to decide its destiny" ([Jaspers 1932] 1970, 185) and "(w)e become ourselves by entering with open eyes into the limit-situations" (179, translation modified).
>
> (Lagerkvist and Andersson 2017, 554f.)

As with ancient Gilgamesh, our encounters with these limit or boundary experiences force us to confront our limitations—first of all, as embodied and thereby vulnerable and ultimately mortal beings. While we, like Gilgamesh, initially seek to ignore or avoid such limit-situations—for example, in hopes of finding an everlasting life—it is only in coming to grips with our mortality and vulnerability that we can take on and exercise our freedom to discern and/or create meaning for our existence.[7]

Gilgamesh suggests that such existential confrontation requires the warrior's greatest *courage*. As Shannon Vallor makes explicit, courage is a central "technomoral virtue," one of twelve she argues are requisite to leading good lives of flourishing in our technological era (2016, 154).

4.2 Existentialism and Virtue Ethics

There are critical overlaps between contemporary existentialism and virtue ethics. Most centrally, Shannon Vallor's *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (2016), is the most extensive and informed exploration of what virtue ethics might look like in our time. Specifically, Vallor's twelve "technomoral" virtues, while squarely rooted in Aristotelian, Confucian, and Buddhist virtue ethics traditions, are carefully redefined vis-à-vis the contexts and demands of our contemporary existence. The list is central here: honesty, self-control, humility, justice, courage, empathy, care, civility, flexibility, perspective, magnanimity, and "technomoral wisdom," that is, "the successful integration of a person's moral habits, knowledge, and virtues in an intelligent, authentic, and expert manner" (2016, 154). At the same time, Vallor's project is also explicitly grounded in existentialism—specifically, the work of Jose Ortega y Gasset who further reiterates the Enlightenment-Romantic view of technology as *emancipating* us for self-cultivation, including our specific existential tasks:

> Ortega y Gasset tells us that our humanity rests entirely upon the "to do" of projected action, and hence "the mission of technology consists in releasing man for the task of being himself."
>
> (Ortega y Gasset 2002, 118, in Vallor 2016, 247)

Here, existentialism thus further intersects with the Enlightenment-Romantic emphasis on technology as emancipatory. More broadly, Vallor's taxonomy of virtues and her more specific application of these in conjunction with social media (ch. 7), surveillance (ch. 8), robots (ch. 9), and human enhancement technologies (ch. 10), thus take us a very long way indeed towards a reasonably complete ethical guide for twenty-first century existence.

5. Courage and Other Virtuous Practices for Good Lives in the Twenty-First Century

I have argued for an account of human beings (a philosophical anthropology) that begins with the fact of our mortality as embodied beings who are likewise deeply relational: Gilgamesh stands as the first and oldest lesson in overcoming existential despair through cultivating the virtues of courage and care in relationships of (deontological) respect and equality. The stress here on embodiment is further enhanced by foregrounding non-dualistic understandings of self and body, and *relational autonomy* as conjoining relationality with autonomy as the ground of modern notions of democratic rights and norms.

Exploring the historical and theological backgrounds of modern technology illustrated how prevailing conceptions of technology as aiming to render us "masters and possessors of nature," who thereby will overcome labor and perhaps even death, rest upon a specific (Augustinian) reading of the 2nd Genesis creation story: this reading enjoins sharp dualisms between soul and body, humanity and nature, male and female, as well as human and technology, as manifested in fears of a Frankenstein and the "techno-femme fatale" of female robots. By contrast, traditional Jewish and early Christian readings of this story foster non-dualistic understandings, for example, of the relationships between male and female and humanity and nature. These understandings cohere with the initial elements of the philosophical anthropology, beginning with embodiment, relational autonomy, care and equality—and are further elaborated in Mark Coeckelbergh's account of the complementarities (rather than opposition) between the Enlightenment and Romanticism. This account stresses the limits of rationality, so as to open up possibilities of "transcendence," of knowledge and experience beyond (but not necessarily in opposition to) reason alone. Within this framework, technology is envisioned as emancipatory.

Lastly, contemporary existentialism reiterates themes of embodiment and mortality, especially by way of Karl Jaspers' conceptions of the limit situation, *existenz*, and *transzendenz*. At the same time, existential interests undergird and intersect with contemporary virtue ethics as elaborated by Shannon Vallor—including the technomoral virtue of courage.

With this philosophical anthropology and ethical toolkit now more fully developed, I offer some final suggestions for pursuing good lives of meaning and flourishing as human beings inextricably interconnected with our technologies. Kant's Enlightenment motto—"Dare to be wise! Have the courage to use your own understanding!" (Kant [1784] 1970, 54)—is a necessary starting point, one that directly intersects with Vallor's exploration of courage and its close relatives, namely, hope, perseverance, and fortitude (2016, 129–132).[8] Following Kant, a first responsibility is the responsibility to *know*. We can think of the pursuit of knowledge and understanding as a practice, as a virtue—one that, Kant makes explicit, requires courage. This begins with our recognition of our mortality—a recognition that proved challenging to even the mighty Gilgamesh. But this also entails knowledge in more than a purely intellectual or abstract sense. Coherent with emphasizing embodied knowledge, here we can note that our existential beginnings often start with a more immediate and *felt* recognition of our mortality. As with Gilgamesh, this may come on the occasion of the death of someone close to us. For many contemporaries, we know—more precisely, as it is said in Norwegian, we know or understand on our bodies (*skjønner på kroppen*)—our mortality when diagnosed with possible cancer or some other life-threatening disease. In these ways, the Kantian motto can be expanded: have the courage to know and to *feel* through embodied knowledge.

Such knowledge is always difficult: but it may be more difficult in the contemporary era, which is deeply marked by the second Cartesian dream—to overcome sickness, perhaps even death. Specifically, from *Neuromancer*, William Gibson's foundational novel of cyberspace (1984) through transhumanism and its philosophical siblings—the foundational assumptions of what we call "the digital era" are shaped precisely by a Cartesian dualism and, in Gibson's case, an explicitly Augustinian contempt for the body (Ess 2011). Swimming against these streams makes the existential project of coming to grips with our mortality that much more difficult.

Socrates would further remind us that such knowledge must begin with self-knowledge (*Phaedrus* 229E; 1914, 221): and, as Foucault argued, such knowledge depends on the *virtue* of *self-care* (*epimelēsthai sautou*), especially as exercised through the technologies of writing (Foucault 1988; Ess 2014). Indeed, from an existential perspective, writing is central to confronting the problems of meaning, beginning with nihilism: "Nihilism . . . always threatens. But the romantic solution is to keep meaning moving through space, to write as if one's very life depended on it—as it does" (Black 2002, 143, in Coeckelbergh 2017a, 62). In an era which seems increasingly bent on short tweets and ever-more emojis as primary forms of expression, the cultivation of the self through careful and reflective writing is all the more urgent.

5.1 Courage, Resistance, and "Hacking" Our Existential Vulnerabilities

Recalling the woman in the Garden and her descendants, from Antigone and Socrates forward: resistance and disobedience are required for the sake of acquiring our emancipation, for achieving and cultivating our own autonomy, maturity, and capacities for independent judgment (*phronēsis*). They are further required as moral heroes from Antigone to Tess Asplund (Crouch 2016) demonstrate: the courage to disobey and resist is vital to emancipation, to opposing the multiple structures and movements that entail our continued, if not enhanced, subordination and obedience (Ess 2019). And coherent with non-dualistic embodiment—such resistance must include our bodies in the streets if it is to be successful (Lim 2018).

Perhaps few of us (as Plato suggests) can rise to the demands and costs of such moral heroism. But resistance in the name of emancipation can—and should—be practiced in more everyday ways as well, again for the sake of sustaining and enhancing the virtues requisite for good lives and flourishing. First of all, from an existential perspective, fully and deeply confronting the realities of our mortality and vulnerabilities requires the courage to resist the multiple ways in which our lives and social practices conspire to silence and ignore mortality—perhaps all the more so as in a digital era that denigrates the body and "meatspace" in the name of vague hopes for digital immortality.

Astrid Hovde (2016) has documented important examples of such courage and resistance in her interviews with the bereaved who first learned of the death of a sibling, close friend, or child through a hasty posting to their Facebook page from someone seeking to express condolences. Certainly, some important positives follow from diverse uses of social media in these experiences. But for six of ten interviewees, multiple aspects of such online grieving and memorializing were increasingly experienced as fake, as more self-interested expressions from persons who paid no attention to them in their offline encounters (Hovde 2016, 101). Not surprisingly, several found that grieving required the embodied co-presence of others—family and close friends—who could hold and comfort them in their deepest moments of sorrow and anguish. The sharp contrast between these online and offline experiences inspired two interviewees ("Sophie" and "Elisabeth") to dramatically reduce their use of Facebook (Hovde 2016, 51–59)—a form of courageous resistance and disobedience to prevailing norms, especially for young people in Norway who are among the most active users of social media. These turns away from social media were, however, necessary in order to confront the deeply existential experiences of loss, and to then take these up, in Jaspers' and Arendt's terms, as limit-situations that open up new possibilities for us (natality)—specifically as they found therein ways of moving into new stages of independence and relationship.

Echoing Mark Coeckelbergh's phrase ("hacking our vulnerability," Coeckelbergh 2017b), such cracking through otherwise prevailing death-denying practices and consensual illusions can be thought of as "existential hacking"—not only for the sake of slipping off one or more forms the digital shackles that prevent us from fully confronting our mortality, but also for the sake of discernment, creativity and meaning-making. Indeed, there are multiple examples of hacking our digital vulnerabilities, precisely in the name of flourishing and good lives. For example, "digital sabbaths"—that is, extended periods of time entirely offline, in order to cultivate significant experiences in both solitary and social ways—appear to be increasingly popular and recommended (e.g., Syvertsen and Enli 2020).

More ambitiously, applying Enlightenment courage for the sake of cultivating our own (relational) autonomy and judgment entails becoming ever-more savvy about our digital technologies: like hackers, by becoming more familiar with how they work, their affordances, as well as their risks and dangers, we are thus better able to circumvent and reshape these technologies in the name of more humane lives and flourishing societies. These directions can entail conscious steps not only to increase the privacy and security of our devices and communications (e.g., by taking the simple steps of encrypting our storage devices and moving to encrypted communication channels such as WhatsApp and Signal) but also to counter the multiple forms of surveillance to which we are increasingly subject (Hildebrandt 2016). We can further participate in the growing number of "hacker spaces" or "maker spaces" in which interested folk with diverse skills and interests come together to learn new skills (from soldering to programming) that allow them to create new devices, whether utilitarian and/or aesthetic. These new creations directly instantiate and echo the "romantic machines" foregrounded by Coeckelbergh as early examples of the coherency between Enlightenment rationalism and romanticism (2017a, 3). They further exemplify Coeckelbergh's attention to the role of crafts and cultivating "skilled practices" in the Romantic effort to reshape and re-appropriate technologies as part of cultivating selves, relationality, and good lives more broadly (2017a, 60, 55).

Certainly, readers will discern still further ways of pursuing such directions, in the name of emancipation and the pursuit of good lives of meaning as embodied and relational beings. While I have argued that we can do well by beginning with the ancient Gilgamesh—as the transformations of existentialism and virtue ethics vis-à-vis our co-evolving technologies make clear, we will inevitably discern and create meaning in new ways as well. *Bon voyage! Bon courage!*

Notes

1. "A" and "B" denote the standard reference to the first and second edition of *Die Kritik der reinen Vernunft* [Critique of Pure Reason], 1781 and 1787, respectively.
2. The oral sources for the subsequent versions of the written epic may be as old as 2700 BCE (Kovacs 1985, xxii). The written sources may be as early as 2000 BCE—well over a millennium before Homer's *Illiad* and *Odyssey* (Kovacs 1985, xxiii).
3. It is also arguable that Kant's account of the free human being is already more relational than the Hobbesian atomistic conception (Hongladarom 2007).
4. Careful interpreters note that the woman does not receive her name—Eve, the mother of all living—until the primal couple are expelled from the Garden (Gen 3.20).
5. Instead of using "versus" I intentionally use "vis-à-vis" (literally, face-to-face) to indicate the whole range of possible relationships between two relata, for example, as equals, as superior/inferior, as complementary, as oppositional, etc. The point is to *not* assume that we already know what these relationships may be: rather, they must be explored, discerned, and justified in turn.
6. A remarkable instantiation of this collocation of "romantic machines" aimed towards emancipation, democracy, and the autonomy requisite for virtue ethics' conceptions of good lives and flourishing is in the project of the IEEE (International Electrical and Electronic Engineers) to establish global standards for "ethically-aligned design" of Artificial and Independent Systems (2019). The first guidelines emphasize (deontological) human autonomy and dignity—and explicitly begin with Aristotle's conception of *eudaimonia* or well-being as its central ethical pillars (2019, 4). This project simultaneously instantiates a further Romantic emphasis—namely, the focus on the importance of *design* of our machines and technologies: see Coeckelbergh 2017a, 56ff.
7. In this (and multiple other ways) Jaspers is deeply influenced by Kant (Thornhill & Miron 2018)—specifically, Kant's central concepts of the *noumena* vis-à-vis the *phenomena* function as a "limiting concept" (*Grenzbegriff*) (Kant [1781, 1787] 1965, 107; A 255/B 310f.). We also saw that this distinction is further central to Coeckelbergh's account of a deeply entangled Enlightenment rationalism and Romanticism.
8. Again, this endorsement of Kant is not naïve regarding the extensive critiques of Enlightenment reason, including his cosmopolitanism, as explicitly aimed against European colonization and imperialism, as nonetheless entangled with racism, etc. The debates here go well beyond the boundaries of this chapter, but in the end, I side with the postcolonial scholar Inder S. Marwah, who explores the interconnections between eighteenth and nineteenth century liberal political thought and imperialism: for all of the deeply

seated problems and complexities, Marwah ends by endorsing Kant's unambiguous recognition of "the inalienable humanity in us all as commanding respect" and thereby equality (2012, 405). In addition, "Kant's moral and political philosophy provides a deep well of resources upon which to draw in arguing for the quality of all cultures" (Marwah 2012, 406).

References

Baier, Annette C. 1987. "Commodious Living." *Synthese* 72, no. 2: 157–185.

Black, John David. 2002. *The Politics of Enchantment: Romanticism, Media, and Cultural Studies*. Waterloo: Wilfrid Laurier University Press.

Boden, Margaret. 2006. *Mind as Machine: A History of Cognitive Science*. Oxford: Clarendon Press.

Buck-Morss, Susan. 2000. "Hegel and Haiti." *Critical Inquiry* 26, no. 4: 821–865.

Coeckelbergh, Mark. 2017a. *New Romantic Cyborgs: Romanticism, Information Technology, and the End of the Machine.* Cambridge, MA: MIT Press.

Coeckelbergh, Mark. 2017b. "Vulnerable Cyborgs Reloaded: From Narrative Technologies for Self-Shaping to Vulnerability Transformations in Modernity and the Anthropocene." Keynote Address: "Precarious Media Life." Sigtuna, Sweden, October 30–November 1.

Consalvo, Mia. 2004. "Borg Babes, Drones, and the Collective: Reading Gender and the Body in Star Trek." *Women's Studies in Communication* 27, no. 2: 177–203.

Crouch, David. 2016. "Woman Who Defied 300 Neo-Nazis at Swedish Rally Speaks of Anger." *The Guardian*, 4 May. https://www.theguardian.com/world/2016/may/04/woman-defied-neo-nazis-sweden-tess-asplund-viral-photograph

Descartes, René. [1637] 1972. "Discourse on Method." In *The Philosophical Works of Descartes*, translated by E. S. Haldane and G. R. T. Ross, Vol. I, 81–130. Cambridge: Cambridge University Press.

Dhawan, Nikita. 2014. "Introduction." In *Decolonizing Enlightenment: Transnational Justice, Human Rights and Democracy in a Postcolonial World*, edited by Nikita Dhawan, 9–16. Verlag Barbara Budrich: Leverkusen, Germany. doi: 10.2307/j.ctvddzsf3

DORS4 (Death Online Research Symposium 4). 2018. University of Hull, UK, August 15–17. http://cc.au.dk/en/research/research-programmes/cultural-transformations/cultures-and-practices-of-death-and-dying/dorn/death-online-research-symposium-4/

Ess, Charles. 1995. "Reading Adam and Eve: Re-Visions of the Myth of Woman's Subordination to Man." In *Violence Against Women and Children: A Christian Theological Sourcebook*, edited by Carol J. Adams and Marie M. Fortune, 92–120. New York: Continuum.

Ess, Charles. 2004. "Beyond *Contemptus Mundi* and Cartesian Dualism: Western Resurrection of the BodySubject and (re)New(ed) Coherencies with Eastern Approaches to Life/Death." In *Philosophie des Todes: Death Philosophy East and West*, edited by Günter Wollfahrt and Hans Georg-Moeller, 15–36. Chora Verlag: Munich.

Ess, Charles. 2010. "The Embodied Self in a Digital Age: Possibilities, Risks, and Prospects for a Pluralistic (democratic/liberal) Future?" *Nordicom Information* 32, no. 2: 105–118.

Ess, Charles. 2011. "Self, Community, and Ethics in Digital Mediatized Worlds." In *Trust and Virtual Worlds: Contemporary Perspectives*, edited by Charles Ess and May Thorseth, 3–30. Oxford: Peter Lang.

Ess, Charles. 2014. "Selfhood, Moral Agency, and the Good Life in Mediatized Worlds? Perspectives from Medium Theory and Philosophy." In *Mediatization of Communication* (vol. 21, *Handbook of Communication Science*), edited by Knut Lundby, 617–640. Berlin: De Gruyter Mouton.

Ess, Charles. 2017."God Out of the Machine? The Politics and Economics of Technological Development." In *Macmillan Interdisciplinary Handbooks: Philosophy,* edited by Anthony Beavers, 83–111. Farmington Hills, MI: Macmillan Reference.

Ess, Charles. 2019. "Ethics and Mediatization: Subjectivity, Judgment (*phronēsis*) and Meta-Theoretical Coherence?" In *Responsibility and Resistance: Ethics in Mediatized Worlds*, edited by Tobias Eberwein, Matthias Karmasin, Friedrich Krotz, and Matthias Rath, 71–90. Berlin: Springer. doi: 10.1007/978-3-658-26212-9_5

Ess, Charles, and Hallvard Fossheim. 2013. Personal Data: Changing Selves, Changing Privacies. In *Digital Enlightenment Forum Yearbook 2013: The Value of Personal Data* , edited by Mireille Hildebrandt, Kieron O'Hara, and Michael Waidner, 40–55. Amsterdam: IOS Amsterdam.

Foucault, Michel. 1988. Technologies of the Self. In *Technologies of the self: A Seminar with Michel Foucault,* edited by Luther H. Martin, Huck Gutman, Patrick H. Hutton, 16–49. Amherst: University of Massachusetts Press.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. London: Penguin.

Hildebrandt, Mireille. 2016. *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Northampton, MA: Elgar.

Hobbes, Thomas. 1651. *Leviathan or the Matter, Forme, & Power of a Common-wealth Eccelsiastical and Civill*. London: Andrew Crooke [public domain].

Hongladarom, Soraj. 2007. "Analysis and Justification of Privacy from a Buddhist Perspective." In *Information Technology Ethics: Cultural Perspectives*, edited by Soraj Hongladarom and Charles Ess, 108–122. Hershey, PA: Idea Group Reference.

Horkheimer, Max, and Theodor W. Adorno, [1947] 2002. *Dialectic of Enlightenment: Philosophical Fragments*, edited by Gunzelin Schmid Noerr, translated by Edmund Jephcott. Stanford: Stanford University Press.

Hovde, Astrid Linnea Løland. 2016. *Grief 2.0: Grieving in an Online World*. MA thesis, Department of Media and Communication, University of Oslo. https://www.duo.uio.no/bitstream/handle/10852/52544/Hovde-Master-2016.pdf?sequence=5

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2019. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html

Jaspers, Karl. [1932] 1970. *Philosophy*, vol. II. Chicago: University of Chicago Press.

Kant, Immanuel. [1781, 1787] 1965. *Immanuel Kant's Critique of Pure Reason.* Translated by Norman Kemp Smith. New York: St. Martin's Press.

Kant, Immanuel. [1784] 1970. "An Answer to the Question: 'What Is Enlightenment?'" Translated by H. B. Nisbet. In *Kant: Political Writings*, edited by Hans Reiss, 54–61. Cambridge: Cambridge University Press.

Kant, Immanuel. [1785] 1993. *Grounding for the Metaphysics of Morals*, 3rd ed. Translated by James W. Ellington. Indianapolis, IN: Hackett.

Kant, Immanuel. [1793] 1968. *Die Religion innerhalb der Grenzen der bloßen Vernunft* [Religion within the Bounds of Reason Alone]. In *Kants Werke: Akademie Textausgabe*, Vol. VI, 1–202. Berlin: Walter de Gruyter.

Kant, Immanuel. [1797] 1964. *The Metaphysical Principles of Virtue*. Translated by J. Ellington. New York: Library of Liberal Arts.

Kovacs, Maureen Gallery. 1985. *The Epic of Gilgamesh*. Stanford, CA: Stanford University Press.

Lagerkvist, Amanda, ed. 2018. *Digital Existence: Ontology, Ethics and Transcendence in Digital Culture.* London: Routledge.

Lagerkvist, Amanda, and Andersson, Yvonne. 2017. "The Grand Interruption: Death Online and Mediated Lifelines of Shared Vulnerability." *Feminist Media Studies* 17, no. 4: 550–564. doi: 10.1080/14680777.2017.1326554

Lim, Merlyna. 2018. "Roots, Routes, Routers: Communications and Media in Contemporary Social Movements." *Journalism and Communication Monographs*, 20, no. 2: 92–136.

Lomborg, Stine. 2012. Negotiating Privacy through Phatic Communication: A Case Study of the Blogging Self. *Philosophy and Technology* 25: 415–434. doi: 10.1007/s13347-011-0018-7.

Marwah, Inder S. 2012. Bridging Nature and Freedom? Kant, Culture, and Cultivation. *Social Theory and Practice*, 38, no. 3: 385–406.

MacIntyre, Alasdair. 1994. *After Virtue: A Study in Moral Theory* (2nd edition). Duckworth, UK: Guilford.

Marx, Karl, and Engels, Friedrich. [1846] 1976. *The German Ideology*. In *Karl Marx and Friedrich Engels: Collected Works,* vol. 5. New York: International Publishers and Progress Publishers.

Natanson, Maurice. 1970. *The Journeying Self: A Study in Philosophy and Social Role*. Reading, MA: Addison-Wesley.

Ortega y Gasset, José. 2002. *Toward a Philosophy of History.* Translated by Helene Weyl. Urbana and Chicago: University of Illinois Press.

Peters, John Durham. 2015. *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. Chicago: University of Chicago Press.

Ruddick, Sara. 1975. Better Sex. In *Philosophy and Sex*, edited by Robert Baker and Frederick Elliston, 280–299. Amherst, NY: Prometheus Books.

Sandars, N. K. 1972. *The Epic of Gilgamesh: An English Version with an Introduction*. New York: Penguin Books.

Shelley, Mary Wollstonecraft. [1818] 2003. *Frankenstein; or, The Modern Prometheus*, edited by Maurice Hindle. Rev. ed. London: Penguin.

Solomon, Robert, ed. 1972. *Phenomenology and Existentialism*. New York: Harper & Row.

Syvertsen, Trine, and Enli, Gunn. 2020. Digital Detox: Media Resistance and the Promise of Authenticity. *Convergence: The International Journal of Research into New Media Technologies*, 26, no. 5–6: 1269–1283. doi: 10.1177/1354856519847325.

Tappolet, Christine. 2014. "Emotions, Reasons and Autonomy." In *Autonomy, Oppression and Gender*, edited by Andrea Veltman and Mark Piper, 163–180. Oxford: Oxford University Press.

Thornhill, Chris, and Miron, Ronny. 2018. "Karl Jaspers," *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2018/entries/jaspers/.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Cambridge, MA: MIT Press.

Veltman, Andrea. 2014. "Autonomy and Oppression at Work." In *Autonomy, Oppression and Gender*, edited by Andrea Veltman and Mark Piper, 280–300. Oxford: Oxford University Press.

Veltman, Andrea, and Mark Piper. 2014. "Introduction." In *Autonomy, Oppression and Gender*, edited by Andrea Veltman and Mark Piper, 1–11. Oxford: Oxford University Press.

Westlund, Andrea. 2014. “Autonomy and Self-Care.” In *Autonomy, Oppression and Gender*, edited by Andrea Veltman and Mark Piper, 181–198. Oxford: Oxford University Press.

Wilson, Robert A., and Lucia Foglia. 2017. “Embodied Cognition.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/.

Winner, Langdon. 1977. “Frankenstein’s Problem.” In *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*, chap. 8, 306–335. Cambridge, MA: MIT Press.

Zuidervaart, Lambert. 2015. Theodor W. Adorno. *The Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2015/entries/adorno/.

CHAPTER 30

WHY CONFUCIANISM MATTERS FOR THE ETHICS OF TECHNOLOGY

PAK-HANG WONG

The idea that Confucianism matters to ethics of *technology* may seem peculiar, as it has long subordinated the interest in science and technology to the pursuit of ethical perfection and thus has undervalued the role of science and technology. However, in the mid-twentieth century, contemporary New Confucianism (re-)affirmed the importance of democracy, science, and technology alongside Confucianism for the future of Chinese culture. The New Confucians do not only argue for the compatibility of Confucianism with modern science and technology, but also for the possible contribution of Confucian values to a more humane development of science and technology (He 2018). Hence, the idea that Confucianism does matter to ethics of technology should not be too surprising.

Indeed, there are recent attempts to introduce Confucian values to ethical analysis of technology (see e.g. Wong 2012; Vallor 2016). These works, however, have not attended sufficiently to one central aspect of Confucianism, namely Ritual ("*Li*"). *Li* is central to Confucian ethics, and it has been suggested that the emphasis on *Li* in Confucian ethics is what distinguishes it from other ethical traditions (see e.g. Fan 2010; Bockover 2012; Stalnaker 2016; Olberding 2015, 2016). Accordingly, any discussion of Confucian ethics for technology remains incomplete without accounting for *Li*. The aim of this chapter, therefore, is to elaborate the concept of Confucian *Li* and discuss its relevance for ethical reflection of technology.

I begin with Joel Kupperman's critique of mainstream analytic ethical theories as being irrelevant and incomplete, and then suggest that his critique also applies to the current ethics of technology. Kupperman's critique usefully reminds us of the ethical importance of styles of interaction and, relatedly, the role of Confucian *Li* in informing and guiding the styles, which have so far escaped the attention of philosophers and ethicists of technology. Hence, I shall elaborate on the idea of Confucian *Li* and examine its role in ethical reflection. After illustrating the idea of Confucian *Li*, I shall discuss different ways in which it is relevant to the ethical analysis of technology. Particularly, by analyzing *Li*'s communicative, formative, and aesthetic functions, I formulate an approach to ethics of technology with an emphasis on *community*, *performance*, and *the aesthetic* and demonstrate how, based on Confucian *Li*, a Confucian ethics of technology may work. In doing so, I hope to have answered the question: *why Confucianism matters in ethics of technology*.

1. Big Moment Ethics, Ethics of Technology, and the Ethical Importance of Style

Joel Kupperman (2002, 2007, 2010) argues for the importance of Confucian ethics by noting a significant gap in mainstream analytic ethical theories. He characterizes mainstream ethical theories as "big moment ethics" that centers on high-stakes ethical decisions in infrequent, one-off situations, which are often presented in a decontextualized manner. An obvious example is the trolley problem, where we are asked to decide whether one should sacrifice one life to save five, but have been provided artificial and/or minimal details of the scenario.[1] Kupperman (2007) argues that the "big moment ethics" is unsatisfactory, as the ethical judgments derived from the decontextualized cases often do not generalize once contextual details are supplied. "Big moment ethics," therefore, is unhelpful in guiding our judgments and behaviors in a contextualized and richly textured ordinary life. More importantly, Kupperman points out that, by focusing on the infrequent, one-off situations, the "big moment ethics" has truncated ethical reflection. It leaves out most of our everyday life as an "ethical free-play zone, in which one can do whatever one likes [and] yields an ethics that does not make demands at all often, and certainly not continuously" (Kupperman 2002, 40). "Big moment ethics" thereby omits ethically significant issues in everyday life that demand a sustained effort, such as a person's style of life, personal relationships, and self-improvement. In short, Kupperman criticizes mainstream ethical theories as *irrelevant* and *incomplete*, that is—the decontextualized examples discussed in mainstream ethical theories offer little guidance for ordinary situations, and they also neglect meaningful ethical questions in everyday life that require on-going reflection by focusing on the rare, one-off cases.

In ethics of technology, although there are discussions focusing on rare, one-off scenarios—for example, existential risks (Bostrom 2002) or debates highly speculative in nature (cf. Nordmann 2007; Nordmann & Rip 2009) that are susceptible to Kupperman's critique—the field has undergone a number of "turns" that seem to have addressed Kupperman's charge to mainstream ethical thought. For example, since "the empirical turn," philosophers and ethicists of technology have paid close attention to how different technologies are *actually* created, how they *actually* work, and how they *in reality* co-shape the self and society with their designers, users, and other related parties (Kroes & Meijers 2000; Brey 2010). Also, "the design turn" (van den Hoven 2008) and "the axiological turn" (Kroes & Meijers 2016) have invited philosophers and ethicists to explicate values in technology and proactively embed them into various technologies to make those technologies conducive to human well-being and to a good society. So construed, the current ethics of technology do attend to the specifics of technology and everyday life and allow a much broader scope of ethical reflection than the "big moment ethics."

For instance, postphenomenology, one of the most elaborated approaches in philosophy and ethics of technology since the empirical turn (see e.g. Ihde 1990; Verbeek 2005; Rosenberger & Verbeek 2015), can be viewed as an answer to Kupperman's critique. Postphenomenology examines and evaluates how various technologies mediate the relations between human beings and the world, and it discusses ways to improve individual and societal well-being through different forms of technological mediation via the design and (everyday) use of technology. For example, Peter-Paul Verbeek (2011, 85–87) refers to obstetric ultrasound, which could confront expectant parents with the dilemma to choose between uncertainty of serious health issues with the unborn, and certainty accompanied by high risk of miscarriage with additional antenatal examinations. He argues that ultrasound re-organizes the expectant parents' moral subjectivity by shaping the ways the fetus can be interpreted and requiring them to decide and act on the (new) information. He also argues that explicitly reflecting on the mediating role of ultrasound opens up the possibility for the expectant parents to actively shape their mediated moral subjectivity by using ultrasound differently; for example, *only* for determining pregnancy due date, or *only* for risk estimation, or even by refraining from ultrasound. In short, Verbeek's postphenomenological approach proposes deliberately using and designing technology to shape human subjectivity and establish oneself as an ethical subject, which is taken to be a continuous (self-)practice. To ethics of technology—at least, to those approaches that take seriously these recent "turns" in philosophy of technology—Kupperman's critique no longer seems applicable.

Yet in an analysis of tele-monitoring systems in Dutch homecare, Ike Kamphof (2017) argues that the postphenomenological approaches have overemphasized the role and power of individuals in shaping human subjectivity through the use and design of technology and have also underplayed the significance of relations between individuals in incorporating (new) technology into practices. She argues that the need to maintain good relations with others (in Kamphof's example, the good relation between caregivers and elderly clients) should inform how a technology is to be used, and can be achieved only by carefully balancing users' feelings, the feelings of others in the relation, and the environment where the technology is being used. Here, Kamphof's argument usefully draws our attention to the fact that a proper use (and design) of technology does not merely amount to a shaping of *oneself* or establishing *oneself* as an ethical subject, but it must include the thoughts and feelings at the recipient end, and thus it is *inevitably* relational.[2]

Although Kamphof has not explicitly formulated her argument in terms of *styles of interaction*, she rightly emphasizes that good relations between individuals (e.g., the caregivers and the elderly clients) are maintained as much by an appropriate way of relating to each other through technology as they are by using (or non-using) technology for suitable ends. For instance, Kamphof (2017, 416–417) observes that the caregivers would reframe their view of a tele-monitoring system from a tool for surveillance and check-up to a device that speaks *for their clients*, and therefore a means to protect *the elderly clients' personal dignity*. They would limit where the system could and should be used in their relationship *with their clients* in both the clients' home and their workplace. By emphasizing the self and subjectivity, postphenomenological approaches have not sufficiently accounted for this style of interaction *with others* through technology.[3] In this respect, Kupperman's critique remains applicable to ethics of technology to the extent that the existing approaches fail to sufficiently integrate people's style of interaction, personal relationships, and self-improvement in the ethical reflection of technology.[4]

If Kupperman's critique remains relevant, his insights on the contribution of Confucian ethics to mainstream ethical thought should also be relevant to the ethical reflection on technology. Before elaborating Kupperman's view in detail, however, it is helpful to explain why Confucian ethics is particularly helpful in foregrounding or capturing the relational dimension of ethics and the on-going nature of ethical reflection which Kupperman deems essential to ethical reflection.

From the Confucian perspective, the notion of personhood is characterized as *relational* and *developmental* (Yu & Fan 2007; Wong 2012).[5] The Confucian notion of person is relational, as Confucians believe that human beings are born into a web of familial and social relationships, and that they can only mature and flourish within such a web by fulfilling the obligations prescribed by their roles and relationships. Roles and relationships, therefore, are necessarily foregrounded in Confucian ethics as they are its normative foundation. Also, the Confucian notion of person is developmental, as Confucians understand personhood to be neither static (i.e., a person is not to be identified by any sets of characteristics) nor given (i.e., human beings are not born as persons). Instead, human beings learn and practice in everyday life to *become* persons. Hence, Confucian ethical cultivation is necessarily an ongoing process that covers every aspect of one's life. Here, philosophers and ethicists of technology can already learn from the Confucian notion of personhood by recognizing the central and essential place of personal roles and relationships in the making of ethical judgments and by reconsidering the significance of the mundane in ethical life (Wong 2012).

Kupperman introduces "naturalness" (or "harmony") as another normative concept that mainstream ethical thought can learn from Confucian ethics. By "naturalness," Kupperman refers to the idea that "the agent is reasonably comfortable with her or his behavior, and there is no conflict between the behavior and what the agent normally is like" (Kupperman 2002, 44). He illustrates the idea of naturalness (of behaviors) with the expression of gratitude: many of us can say "thank you" at ease in return for a favor done, but children may have difficulties in their expression of gratitude; that is, children may forget to do so as they get overwhelmed by the favor. They may be confused and hesitate to say "thank you," or they may simply be rude and thus have to be reminded. In the case of children who are not at ease and fluent in expressing gratitude, even if they *do* say "thank you," their behaviors are not natural (or harmonized), and the unnaturalness of behaviors demonstrates something amiss *ethically*. As Kupperman argues, people's style, that is, *how* something is done and said, presents and reveals their attitudes and who they are, which, in turn, is essential in building and maintaining personal relationships (Kupperman 2002, 2007). So the children who reluctantly say "thank you" may have *said* "thank you," but their style of interaction has failed to convey thankfulness or show themselves *to others* that they are a gracious person. Interestingly, the ethical import of styles of interaction have also been asserted by enactivist philosophers and cognitive scientists, who point out that "different styles of interaction, with their varying affective overtones, will make an ethical difference, in the sense that they will modulate the ethical coloring of any given situation to which the categories of ethical description or appraisal may apply" (Colombetti & Torrance 2009, 520; also, see Hutton 2006).

The Confucian ideal of naturalness, therefore, compels us to consider the ethical import of not only *what* we should do and say, but also *how* we should do and say them—or, as Kongzi remarks on filial piety in *The Analects* 2.7 and 2.8:

> The Master said, "Nowadays 'filial' means simply being able to provide one's parents with nourishment. But even dogs and horses are provided with nourishment. If you are not respectful, wherein lies the difference?" (Slingerland 2003, 10).
>
> The Master said, "It is the demeanor [of filial piety] that is difficult. If there is work to be done, disciples shoulder the burden, and when wine and food are served, elders are given precedence, but surely filial piety consists of more than this" (Slingerland 2003, 10).

It is important to act and speak with an appropriate attitude—even when *what* we do and say are already the morally *right* things to do and say, for example providing for parents, shouldering teacher's burden of work, or giving elders precedence, as our attitudes and our self are expressed and revealed by *how* we do and say the right things.

In short, Confucian ethics recommends a close look at people's style of interaction, for it communicates people's attitudes (about others) and shows themselves to others, which are essential in ethically fruitful connections with others.[6] Yet what does the shift to the style of interaction as recommended by Confucian ethics mean to the ethical reflection on technology? Or, simply, from the Confucian perspective, how can styles of interaction be introduced to ethics of technology? To answer these questions, it is essential to first discuss what guides people's style of interaction. For Confucians, the answer is ritual ("*Li*"): it is *Li* that informs *what* and *how* people should do and say in different personal and social circumstances.

2. A Primer on Confucian Ritual (“*Li*”)

The Analects 12.1 writes, “[r]estraining yourself and returning to the rites [‘*Li*’] constitutes Goodness [‘*Ren*’]” (Slingerland 2003, 125); Confucian *Li*, often translated as “ritual,” “rites,” or “etiquette,” assumes an essential role in Confucian ethics as a *normative* standard for judgment and behavior, and it also informs and guides people’s style of interaction.[7]

In Confucian philosophy,[8] *Li* refers to both ceremonial and formal rituals (e.g., sacrificial offerings, burial ceremonies, and mourning practices) and behavioral patterns for everyday encounters. Accordingly, Confucian *Li* is not a set of *abstract* normative principles, but a collection of *substantive* normative instructions that informs and guides people’s judgment and behavior. Some examples from *The Analects* should be illustrative of the substantive requirements it prescribes:

> When called on by his lord to receive a guest, his countenance would become alert and serious, and he would hasten his steps. When he saluted those in attendance beside him—extending his clasped hands to the left or right, as their position required—his robes remained perfectly arrayed, both front and back. Hastening forward, he moved smoothly, as though gliding upon wings. Once the guest had left, he would always return to report, ‘The guest is no longer looking back.’
>
> (*The Analects* 10.3, in Slingerland 2003, 99)
>
> The gentleman did not use reddish-black or maroon for the trim of his garment, nor did he use red or purple for his informal dress. In the summer, he wore a single layer of linen or hemp but always put on an outer garment before going out. With a black upper garment he would wear a lambskin robe; with a white upper garment he would wear a fawn-skin robe; and with a yellow upper garment he would wear a fox-fur robe. His informal fur robe was long, but the right sleeve was short. He required that his nightgown be knee-length. He wore thick fox and badger furs when at home. Except when he was in mourning, he never went anywhere without having all of his sash ornaments properly displayed. With the exception of his one-piece ceremonial skirts, his lower garments were always cut and hemmed. He did not wear [black] lambskin robes or dark caps on condolence visits. On the day of the ‘Auspicious Moon,’ he would always put on his [black] court attire and present himself at court.
>
> (*The Analects* 10.6, in Slingerland 2003, 100–102)
>
> “He [i.e., Kongzi] would not sit unless his mat was straight (‘*Zheng*’).
>
> (*The Analects* 10.12, in Slingerland 2003, 105)

As these examples in *The Analects* demonstrate, Confucian *Li* ranges from the norms for formal occasions (e.g., receiving guests) to the patterns of behaviors in everyday life (e.g., a person’s clothing and posture), and it prescribes appropriate responses and behaviors to people, with reference to their role(s) and relations with others, in specific social circumstances. It is useful to emphasize that the instructions, which involve

Kongzi as an exemplar, do not only advise *what* is to be done and said but document in minute detail *how* they are to be completed. It is also important to note that, while the instructions in Confucian *Li* appear to be extremely rigid, Confucian ethics does have room for (reflective) non-observance and exceptions to it (Li 2007; Kim 2009, 2010). In fact, Kongzi is described in *The Analects* as "entirely free of four faults: arbitrariness, inflexibility, rigidity, and selfishness" (9.4, in Slingerland 2003, 87). Indeed, the need for flexibility and fluidity should be obvious, as Confucian *Li* depends on people's roles and their relations with the interacting partners as well as the social circumstances where the interaction takes place, which are contextual and finely textured. Hence, personalization and improvisation of *Li* are required for any successful performance (Ames 2002).

In the discussion of the ethical importance of Confucian *Li*, three lines of argument can be discerned. The first line of argument focuses on the *communicative* function of *Li*. For instance, Chenyang Li (2007) conceptualizes *Li* as "cultural grammar" for personal and social interaction within a community. He points out that, like linguistic communication, which is based on languages and their grammatical rules, personal and social interaction takes place against a background of values and is governed by norms of interaction. Accordingly, *Li* serves as a public, shared and comprehensible medium to interpret people's responses and behaviors in various social circumstances. Moreover, since *Li* is passed down from generation to generation, it embodies the values of the tradition and provides a normative standard in accordance with *that* tradition. Successful performance of *Li*, therefore, expresses the values of a community and its tradition, and those who belong to that community, or who are familiar with that tradition, can grasp the meaning (and embodied values) of the performed *Li*. It is in this sense, Mary Bockover (2012) argues, that Confucian *Li* can be viewed as a culturally-specific "body language."

The *ethical* dimension of Confucian *Li*'s communicative function is best described in Kelly Epley's argument for the role of *Li* in caring (Epley 2015). She rightly points out that expressions of need and care are not isolated from social conventions and communal standards of manners. In effect, social conventions and manners play a constitutive role in comprehending need and realizing care. Imagine a person who fails to attend to another person's need because their expressions of need are different; for example, a community where requests for help must be *explicitly* stated (Community A) versus a community that does not require or encourage its members to explicitly request help (Community B). A person from Community A may fail to offer help to a person from Community B even when the latter is clearly in need of help but has not requested it explicitly, and this is the result of their different expressions of need.

Relatedly, a person who is provided care by other people may not be sufficiently cared for when there is a mismatch of the expressions of care. It could be so when the person does not recognize the care provided by others as caring because care is expressed differently in the person's community. Imagine, for example, a case in which the family members conceal the cancer diagnosis from the patient. In an individual-oriented community, this act could be viewed as dismissive and uncaring to the patients, because it ignores their need for truth; however, the same act could be viewed as an expression and act of care in a family-oriented community, where familial values often serve as a reliable source of patients' values and preferences.[9] From a Confucian perspective, in effect, the act of lying is prescribed by the role of being a member of family, which requires them to shoulder the fear, stress, and suffering of the patients (Zhao 2014). It is, therefore, essential to have an understanding of *how* care is expressed and enacted in the community in order to *properly* care. In short, *Li*[10] is ethically significant as a shared resource for understanding and interpreting need and care—or, for that matter, other important shared values as well.

Indeed, *Li* is essential in creating a community of care where the members can recognize the needs of each other and respond appropriately.[11] As Ana S. Iltis (2012, 21–23) has argued, rituals create and shape the social reality of ritual participants and observers by establishing and reinforcing their expectations, relationships, and roles. Hence, knowing rituals means knowing what to expect from others and what others expect from one, and it also means knowing how one is related to others and what *role obligations* one has. A failed ritual performance, therefore, can be seen at once as a communication, social, and *ethical* failure. Here, it is important to reiterate that *Li*—or, social conventions and manners—does not only prescribe *what* a person should do and say, but also *how* it should be done and said, and that both *what* and *how* things are done and said are essential in understanding and interpretation of people's responses and behaviors.[12]

The second line of argument for Confucian *Li* is based on its *formative* function, that is, the practice and performance of *Li* as essential to individual and societal flourishing. Here, Xunzi's description of the formative function of *Li* is instructive:

> Ritual ["*Li*"] cuts off what is too long and extends what is too short. It subtracts from what is excessive and adds to what is insufficient. It achieves proper form for love and respect, and it brings to perfection the beauty of carrying out *yi* ["righteousness"].
>
> (*Xunzi*, chapter 19, in Hutton 2014, 209)

Being concerned with human being's natural inclination towards selfishness, Xunzi argues that *Li* is essential to tame our (excessive) desires and heighten our (deficient) ethical feelings by prescribing appropriate emotional responses and behaviors for various circumstances. It is through practicing and performing *Li* that people become accustomed to the right emotional responses and behaviors, thereby transforming their dispositions (Sung 2012; Olberding 2015, 2016; Stalnaker 2016).[13]

The importance of the bodily-performative dimension of *Li* deserves to be emphasized. As bodily practice and performance, Confucian *Li* must describe *how* it is to be executed to avoid being vacuous.[14] Moreover, the bodily-performative dimension of Confucian *Li* allows people to internalize norms and values, and enables them to react ethically to different situations in spontaneity. This is crucial to individual ethical life because many of our everyday ethical judgments and behaviors are pre-reflective and influenced by the situation (Olberding 2016; also, see Hutton 2006; Slingerland 2011; Seok 2012). It has also been suggested that the practice and performance of Confucian *Li* creates an "as if" space of moral rehearsal, in which people's dispositions are trained and refined (Puett 2015). According to this understanding of *Li*, the bodily-performative dimension is also essential because it is through the (re-)enactment of critical events in the "as if" space, that individuals acquire emotional and physiological experiences and learn to modulate them. The (re-)enactment, therefore, has to include minute details of the critical events in order to fulfil the purpose of training and refinement.

Finally, there is also an *aesthetic* dimension in Confucian *Li* as illustrated in *The Analects* (e.g., 10.6) and in *Xunzi*:

> If your exertions of blood, *qi*, intention, and thought accord with ritual, they will be ordered and effective. If they do not accord with ritual, they will be disorderly and unproductive. If your meals, clothing, dwelling, and activities accord with ritual, they will be congenial and well-regulated. If they do not accord with ritual, you will encounter dangers and illnesses. If your countenance, bearing, movements, and stride accord with ritual, they will be graceful. If they do not accord with ritual, they will be barbaric, obtuse, perverse, vulgar, and unruly.
>
> (*Xunzi*, chapter 2, in Hutton 2014, 10)

When one acts and speaks with Confucian *Li* (i.e., one speaks with appropriate style), one's behavior will be "congenial and well-regulated" and "graceful"—or, more generally, beautiful. Olberding (2015, 2016) explains the ethical and social implications of the *beautification* function of Confucian *Li* by drawing attention to the power of positive aesthetic properties to mitigate pre-reflective, negative impressions that arise from "ugliness" (or incivility) of behaviors and social environments. By conforming to Confucian *Li*, that is, a communal standard of appropriate emotional responses and behaviors, one beautifies one's emotional response and behaviors by making them more pleasant and agreeable, thereby reducing the potential for conflict and encouraging social cooperation.[15] Or, as Yuriko Saito astutely notes, "[t]he aesthetic appeal of an elegant body movement thus is not for the sake of aesthetic effect alone but more importantly a sensuous display of one's other-regarding considerations" (Saito 2017, 211).

To summarize, the aim of this section is to introduce a practicable idea of Confucian *Li* and illustrate its relation to style of interaction. Briefly, Confucian *Li* prescribes *what* a person should do and say, and *how* they should do and say them in accordance with their role(s) and relation(s) with the interacting partners and with the circumstance one finds oneself in. To Confucian ethics, *Li* and the styles of interaction prescribed by Confucian *Li* are essential because they enable people in the community, or those who share a tradition, to communicate meaning and values appropriately. At the same time, practice and performance of Confucian *Li*, understood as a bodily activity, allow individuals to refine and modulate their (pre-reflective) sensibilities of others and the environment. Finally, Confucian *Li* also accounts for the power of aesthetic properties in ethical and social realms.

3. From Rituals ("*Li*") and Technology to Ritualizing Technology

Based on Confucian *Li*, we can now rethink ethical reflection on technology. In this section, I shall describe what the communicative, formative, and aesthetic functions of Confucian *Li* emphasize in the ethical analysis of technology. In doing so, I articulate what Confucian ethics can contribute to ethics of technology, namely a different approach to ethical analysis of technology that focuses on community, performance, and the aesthetic.

The communicative function of *Li* asserts that the styles of interaction prescribed by one's role, social conventions, and manners are parts of a shared medium of meaning and values in a community, and thus thinking with *Li* requires us to consider *how* meaning and values are expressed and revealed through a particular style of interaction at a specific social circumstance. For ethical analysis of technology, this shift to *Li* necessitates an examination not only of *what* values are embedded in technology, but *how* these values are, or can be, manifested through the use of technology and in technologically-mediated interaction. At the same time, this shift to *Li* also implies that we need to consider (1) the recipients, who comprehend and interpret the values expressed and revealed by the use of technology and in technologically-mediated interaction, and (2) the existing styles of interaction, social conventions, and manners in a community, which provide the common ground of understanding and interpretation of need and care as well as other important shared values. Accordingly, a Confucian ethical analysis of technology has to be both (1) relational and (2) communal.

Here, Kamphof's (2017) discussion of how caregivers adopt the tele-monitoring system is instructive. She documents how caregivers use motion sensors in different ways that re-articulate the meanings of privacy *for* and *with* the elderly clients; and, in doing so, the caregivers can respect their clients' privacy while using the tele-monitoring system. For example, the caregivers see and use the tele-monitoring system as a means to protect the clients' dignity by having the devices to speak for them about their undignified problems, and thereby easing the difficulty of conversation with the clients; or, to avoid clients having the feeling of always "being watched," the caregivers would ignore the data they considered meaningless and refrain from communicating *all* information to their clients, thereby making a personal space for them. Kamphof notices that the caregivers' concern is not only about the privacy—or the lack thereof—in the system per se but also about *how* the value of privacy and a good caregiver-patient relation are, or can be, realized in its use with the elderly clients. Kamphof's discussion is illustrative of the importance of *how* to realize values of technology and the relational dimension in the ethical analysis of technology.

From a Confucian perspective, however, what Kamphof's analysis has still missed is the communal dimension for understanding and interpreting the caregivers' use of the system. More specifically, how—or, through which style of interaction—care is expressed and revealed in *that* community, how the introduction of the tele-monitoring system enhances or interferes with the caregivers' original style of interaction, and whether the elderly clients understand the altered style of interaction as care and why. Surely, if a good caregiver-patient relation has been maintained after the introduction of the tele-monitoring system, the elderly clients will see the new style of interaction, as it has been altered by the technology, as care. However, the major insight from the above discussion of Confucian *Li* is that philosophers, ethicists, and technology designers are still in need of a *normative* standard to think through whether, and to what extent, the *different* styles of interaction introduced by the technology are appropriate or not; and, the Confucian ethics of technology answers these questions with reference to *Li*.

What is also missed is the opportunity to use *Li*—that is, the shared medium of meaning and values—to improve technology use and technologically mediated interaction by reproducing or extending the *already* appropriate style of interaction in the design and use of technology.[16] Darian Meacham and Matthew Studley (2017) argue that *robotic expressions* of care, in terms of a robot's gestures, movements, and articulations, are sufficient for caring relations between robots and patients. They maintain that since what matters to patients in a care environment is the caregivers' expressive behaviors but not their mental states, one cannot reject the possibility of caring relations between robots and patients by reason of robots' lacking mental states. More positively, they claim that robots which consistently respond to patients with appropriate expressive and reciprocal movements satisfy basic elements in care; that is, attentiveness, competence, and responsiveness.[17] Here, while the strong claim that robotic expressions of care *constitute* care remains controversial, a weaker claim that robotic expressions of care *communicate* care should be uncontroversial.

From a Confucian perspective, if the robotic expressions do reflect an appropriate style of interaction, which caregivers should have towards the recipients of care, and the patients do view them as communicating care, the idea of "robotic care" will be considered ethically acceptable. Yet it should be reminded that the robotic expressions must *fit* the existing styles of interaction, social conventions, and manners in a community for them to be viewed by the patients (and the caregivers) as care. In this respect, the design and assessment of robotic expressions should be based on the *Li* of care in the community.

In other words, a Confucian approach to ethics of technology proceeds with the existing styles of interaction, social conventions, and manners, and views them as *a* normative basis to evaluate the changes in behaviors and interactions as a result of the use of technology and technological mediation. For example, it may find social media platforms to be ethically problematic, because conventional norms of communication are easily breakable by such platforms due to their design features, which can render the expression and comprehension of meaning and values in a community unstable and ineffective (Wong 2013). Consider the phenomenon of live-reporting overheard conversations on social media platforms; for example, live-tweeting on Twitter. Live-reporting overheard conversations is ethically problematic because it violates conventional norms of communication: where discussing overheard conversations with family and friends *privately* may be innocuous, broadcasting them to the world *publicly* is inappropriate. More important, it is social media platforms like Twitter that blur the boundaries between the public and private spheres and make it more difficult for people to follow the norms of communication. In this way, the Confucian approach can critique social media platforms as running afoul of a normative standard of *Li* for communication.

Alternatively, the Confucian approach also grants that the existing styles of interaction, social conventions, and manners can be employed to improve technology design and use by offering a common medium of meaning and values for designers, users, and recipients, as in the case of robotic care discussed above. To summarize, referring to *Li*'s communicative function, there are two different but related roles for Confucian *Li* in ethical analysis of technology, namely a *normative standard* for ethical analysis and a *normative resource* for devising technology design and use.

The formative function of Confucian *Li* aims at refining and modulating our emotional and physiological experiences, and thereby honing individuals' pre-reflective responses to everyday ethical encounters. Moreover, the refinement and modulation of experiences are to be achieved through bodily practice and performance. Here, Confucian ethics calls for a return to the role of body in ethical development of persons and communities, and thus connects it to the recent research on embodied cognition (Seok 2012; Ott 2017). In a similar vein, ethical analysis of technology should be more receptive to the bodily influences of technology, particularly the possibility of structuring bodily movements through technology design and use (see e.g. Tuuri et al. 2017; Parviainen & Pirhonen 2017) and the affective influences from different technologies (see e.g. Slaby 2016). Technology invites specific bodily and affective interaction with it, just as Confucian rituals of mourning prescribe specific bodily performances and attitudes to members of the deceased's family and other mourners. For example, a laptop computer with keyboard requires its users to *type with their hands* to interact with it, whereas a voice-controlled device requires its user to *speak with their voices* to interact with it. Similarly, people can get *excited* by video games or be *empathic* interacting with (social) robots. Our daily interaction with technology unwittingly induces different emotional and physiological experiences, and these experiences are the basis of our responses to everyday ethical encounters. A Confucian approach to ethics of technology, therefore, should draw our attention to the *bodily* and *affective* impacts of technology with reference to *Li* in a community and its tradition. It may even warrant *proactively* shaping individuals' bodily and emotional states in accordance with Confucian *Li* through the use of technology and in technologically mediated interaction.

Interestingly, Kristina Niedderer (2007, 2014) has advanced the idea of Mindful Design and illustrated the possibility of raising users' attentiveness to the relational, social, and environmental consequences of their actions through the design of objects. For example, she contemplates a design of mobile phone that "shouts back" at its users should they be talking too loudly in public places, thereby alerting users to the disrupting impacts they have on others around them and leading them to adjust the level of their voice (Niedderer 2014). Niedderer's Mindful Design approach converges with Confucian ethics' concerns over people's inappropriate emotional responses and behaviors as a result of their inattentiveness to appropriate styles of interaction, social conventions, and manners for the situations—or, from a Confucian perspective, their inattentiveness to *Li*. In line with Niedderer's Mindful Design, we can imagine the Confucian ethics of technology to advocate designing technology that enables people to be more attentive to the appropriateness of their behaviors with reference to Confucian *Li*. As Eric Hutton (2006) notes, Xunzi has long recognized the role of situational and material factors in self-cultivation:

> Thus, the mourning garments and the sounds of weeping make people's hearts sad. To strap on armor, don a helmet, and sing in the ranks make people's hearts emboldened. Dissolute customs and the tunes of Zheng and Wei make people's hearts licentious. Putting on the ceremonial belt, robes, and cap, and dancing the Shao and singing the Wu make people's hearts invigorated.
>
> (*Xunzi*, chapter 20, cited in Hutton 2006, 44–45)[18]

Recall Xunzi's emphasis on appropriate emotional responses and behaviors. In describing the influences of different situational and material factors on people's "hearts," that is, emotions and/or attitudes, Xunzi highlights the need to consider the situational and material factors in self-cultivation.[19] In fact, Xunzi explicitly endorses using situational factors and material objects in aiding ritual performance, as he states that "fine ornaments and coarse materials, music and weeping, happiness and sorrow—these things are opposites, but ritual makes use of them all, employing them and alternating them at the appropriate times" (*Xunzi*, chapter 19, Hutton 2014, 209). For Confucians, therefore, technology can be "ritualized" to support people's ethical development.

Finally, the aesthetic dimension of Confucian *Li* should also draw attention to the ideal of "beauty" in the ethical analysis of technology. The aesthetic dimension of technology, which broadly refers to the formal, expressive, and sensual properties of technology, is underexplored in ethics of technology, as the focus is primarily on (1) the impacts of (un)intended functions of technology and (2) the moral or political values embedded in technology. From a Confucian perspective, however, the ethical and the aesthetic are intertwined, that is, positive aesthetic features are considered to be *ethically* desirable. Aesthetically pleasing technological design and use are ethically significant because they can reduce potential friction for individuals and in their relationships. However, aesthetic pleasure is also subjected to ethical consideration, and thus is not independent from ethics.

Here, Benjamin Grosser's project *Facebook Demetricator* is instructive (see Grosser 2014). Grosser develops a web browser extension allowing users to hide all the metrics on Facebook and examines users' experience *without* the numeric meters on the social media platforms. Reflecting on the users' feedback, Grosser observes that the display of numeric metrics on Facebook has significant impacts on users' self-understanding and their interaction with others on the platform, as users interpret themselves through the numbers and structure how, when, and with whom to interact based on the numbers. Certainly, Facebook and other social media platforms that display numeric metrics can be critiqued for their *intention* to use these metrics to seduce, manipulate, or abuse their users. Still, a different critique based on the *sensual property* of displaying the numeric metrics is just as important insofar as the sensual property affects their users' decisions and behaviors. So construed, ethics of technology should also account for the aesthetic impacts of technology. Confucian ethics, by foregrounding the link between the ethical and the aesthetic, adds another layer to the ethical analysis of technology, namely the *aesthetic* features in technology design and use. These aesthetic features, of course, are important to the extent that they make technology more pleasant and agreeable in relations and for the community (see e.g. Pols 2017).

4. Conclusion

This chapter aims to explore the contribution of Confucian ethics to the ethical reflection of technology. I propose that Confucian *Li*, with its emphasis on community, performance, and the aesthetic, can enrich the current ethics of technology. As an embodiment of communal and traditional values, Confucian *Li* can be used as a normative standard for ethical analysis of technology, or it can be used to inform the design and use of technology—or, more proactively, to ritualize technology such that it can serve to guide users' and the society's responses with reference to *Li*. However, there are a number of open issues that need to be addressed before the Confucian approach to ethics of technology is fully developed.[20] Before ending this chapter, I shall present some theoretical and ethical challenges related to the communicative and formative functions of *Li* that form the basis of the Confucian approach, as they are helpful to indicate directions for future research.

There are two potential challenges related to the communicative function of *Li*. The first concerns the *normative basis* of *Li* and the second, the possibility of changes in *Li*. So far, I have bracketed the debate on the normative basis of *Li*, and assumed that it *is*, and *should be*, the medium of meaning and values for a community and in a tradition. However, one may object to *Li*, that is, the existing styles of interaction, social conventions, and manners, as a justified *normative* standard by questioning its source of its normativity. Fortunately, there is a recent revival of interest in ethical import of manners and etiquettes that can provide resources to justify the normativity of *Li* (see e.g. Stohr 2011; McPherson 2018). More specifically to the Confucian approach, one—especially those who are non-Confucian—can reasonably question whether and why Confucian *Li* should be *the* normative ground for communicating meaning and values. The answer to this question requires a more careful explication of Confucian *Li* and a detailed comparative analysis of Confucian *Li* with rituals in different communities and traditions, but I should already caution against assuming Confucian *Li* as the only *universal* normative standard available.

This brings us to the second challenge, namely whether and when Confucian *Li* can be altered. The possibility of change is especially important for ethics of technology, as new technology is often "disruptive." Consider the famous Collingridge dilemma, which states that "during [the] early stages of [the development of technology], when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow" (Collingridge 1980, 19). Since Confucian *Li* is based on tradition, it may *not* be able to anticipate and respond to the *novel* consequences that are non-existent in the tradition. Accordingly, the Confucian approach could be dismissive to new practices and relationships mediated by new technology, as it has no normative resources to account for them. The criticism, therefore, is that the Confucian approach will be conservative if it is difficult, or even impossible, for *Li* to be adapted to unprecedented scenarios. This is a criticism the Confucian ought to take seriously; however, I think, there is *no* in-principle reason to reject the possibility of ritual reinterpretation and innovation (see e.g. Neville 2015). Still, in order for the Confucian approach to be defensible, mechanisms for changes in Confucian *Li* ought to be adequately articulated.

In relation to the formative function of Confucian *Li*, a potential concern is the boundary between ethically permissible and impermissible refinement and modulation of emotional and physiological experiences through *Li*. Similar concerns have been raised against nudge theory (Thaler and Sunstein 2008), which aims to influence the individual's decisions and behaviors by (re-)structuring their choice architecture. Particularly, the idea of nudges has been criticized as coercive and undermining autonomy (see e.g. Hausman & Welch 2010). Similarly, Brett Frischmann and Evan Selinger (2018) have described the danger of techno-social engineering of humans, which may strip us of our humanity and turn us into predictable and programmable objects. Here, I think, the Confucian approach affords a different response to this type of concern. First, the perfectionist elements in Confucianism allow it to justify specific refinement and modulation of emotional and physiological experiences with reference to substantive Confucian values (see e.g. Wong 2012; Huang 2015; Sarkissian 2017). Indeed, the grounds for the refinement and modulation ought to be justified by their contribution to human and social flourishing, but not to other *instrumental* ends. Second, Confucian ethics' emphasis on self-cultivation should reject a wholesale techno-social engineering that deprives people's opportunity to cultivate themselves. Yet it remains essential for the Confucian approach to articulate a clearer account of why refinement and modulation of experiences are essential to ethical lives of individuals and which Confucian values can be used in grounding the refinement and modulation.

There is much work to be done to fully articulate a Confucian approach to ethics of technology. So this chapter should only be viewed as a modest attempt to introduce the idea of Confucian *Li* to the ethical analysis of technology. Yet as *community*, *performance*, and *the aesthetic* only have a minor role in the ethical analysis of technology, the

Confucian approach may supplement the current ethics of technology by recovering their importance.

Notes

1. The trolley problem has generated an enormous scholarly discussion, and it is not my intention to discuss it (and other similar ethical dilemmas) in this chapter. The intention is to point out, as Kupperman also does, that mainstream analytic ethical theories often refer to decontextualized cases that are highly unlikely to be encountered by people in their everyday life. For a recent overview of the trolley problem, see e.g. Kamm (2015).
2. Unless, of course, the consequences arise from the use (and design) of technology that is entirely personal. Yet even then it is questionable whether the person who uses this "purely" personal technology can avoid the consideration of others, as their interaction with others may have been altered by the "purely" personal technology.
3. For a defense of the postphenomenological approach from Kamphof's critique, see Sharon (2017). It is useful to point out that Sharon does not reject Kamphof's focus on personal relations, but argues that it offers a supplement but not an alternative to the postphenomenological approach. In this sense, Sharon too acknowledges an emphasis on the role and power of individuals in existing postphenomenological approaches.
4. Here, one may argue that the approaches to ethics of technology inspired by Aristotelian virtue ethics do include personal relationships in their ethical reflection, e.g. Vallor (2016); and, thus even if Kupperman's critique applies to postphenomenological approaches, it does not apply to them. Kupperman's response to this objection comes in two parts: first, he notes that Aristotelian virtue ethics has in fact paid little attention to the style of interaction, understood as the expressions of attitudes and behaviors for specific scenarios (Kupperman 2002); and, second, Kupperman (2004) argues that Aristotelian virtue ethics views ethical decision as a one-person game but not a communal, multi-person game, and thus does not sufficiently capture the relational nature of ethical decision.
5. The Confucian notion of personhood is also characterized as virtue-based. For a detailed discussion of the Confucian notion of personhood and its implication for ethics of technology, see Wong (2012).
6. I believe the aim of Kupperman's critique is to foreground the "hows," which have mostly been ignored in analytic ethical theories. So it is important to note that his critique does not entail a rejection of the "whats" in ethical reflection.
7. The normative priority of *Li* in relation to *Ren*, often translated as "humanity," "goodness," "benevolence," remains a subject of intense discussion. See e.g. Li (2007). I shall not settle the priority between *Li* and *Ren* in this chapter, as an answer to this question has little relevance to the current discussion.
8. My discussion of Confucian *Li* refers primarily to *The Analects* and *Xunzi*, which are considered to be the key texts for understanding the idea of *Li* in (early-) Confucianism (Radice 2017). I should point out that this section is *not* intended to be an exegesis or critical (historical-) textual study of the two texts. The modest aim of this section is to introduce a "workable" idea of *Li* that can enrich ethical analysis of technology.
9. This example is adapted from Zhao (2014).
10. In the example, *Li* is understood broadly as the behaviors, e.g. lying, and attitudes prescribed by the role(s) people occupy without specifying *exactly* what ought to be

performed. *Li* can also be understood narrowly to describe the *exact* behaviors and attitudes to be used in communicating meaning and values. These behaviors and attitudes are often *culturally* and *communally* defined; for example, the performance of "greeting," as an expression and act of friendliness, varies in different cultures and communities.

[11] For a discussion on the community-forming and communal bonding potential of Confucian *Li*, see Bockover (2012).

[12] Buss and Calhoun each offer a similar argument for the ethical importance of manners in terms of their expressive function, see Buss (1999) and Calhoun (2000).

[13] Olberding's discussion of Xunzi's defense of ritual mourning against Zhuangzi's critique offers an instructive example for the working of Confucian *Li*, see Olberding (2015).

[14] Here, a comparison with acquisition of (bodily) skills should be useful. For example, consider learning how to play tennis. It is not sufficient to learn only the rules of the game and the techniques and strategies available; one must also learn how to execute those techniques and strategies. Moreover, tennis players improve their game by honing and refining *the ways* they play; that is, their gesture, strokes, etc. Also, see Stalnaker (2016) for his comparison of ritual with music and cooking.

[15] Also see Kim (2012) for an exposition of Xunzi's view on the function of *Li* in relation to the acquisition of civic virtues.

[16] A similar point has also been made by Pols (2017) in her commentary on Kamphof's analysis without referring to style of interaction, social conventions, or manners but including practices such as " 'being watched' and hence 'looked after,' " "say good-night," etc.

[17] Meacham and Studley (2017, 99) refer to the four "ethical elements of care" presented in Tronto (2005)—i.e., attentiveness, competence, responsiveness, and responsibility—and acknowledge the difficulty for robotic care to satisfy the element of (genuine) responsibility.

[18] Also, see *The Analects* 10.6 (Slingerland 2003, 100–102), where Kongzi's attire is considered as an inherent part of ritual performance.

[19] For further research on the Confucian responses to the situational and material influences on self-cultivation, see Slingerland (2011) and Sarkissian (2017).

[20] I have not included the issues related to the esthetic function of *Li* in the concluding section *not* because I view them to be insignificant, but rather because the relations between the ethical and the aesthetic deserve serious investigation that includes fundamental questions concerning the nature of aesthetic value, the functions of aesthetic activities, etc., which I am unable to account for in this chapter. See Mullis (2007) for a recent attempt to answer these questions from a Confucian perspective.

References

Ames, Roger T. 2002. "Observing Ritual 'Propriety (*Li*)' as Focusing the 'Familiar' in the Affairs of the Day." *Dao* 1, no. 2: 143–156.

Bockover, Mary. 2012. "Confucian Ritual as Body Language of Self, Society, and Spirit." *Sophia* 51, no. 2: 177–194.

Bostrom, Nick. 2002. "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards." *Journal of Evolution & Technology* 9, no. 1. https://www.jetpress.org/volume9/risks.html

Brey, Philip. 2010. "Philosophy of Technology after the Empirical Turn." *Techné* 14, no. 1: 36–48.

Buss, Sarah. 1999. "Appearing Respectful: the Moral Significance of Manners." *Ethics* 109, no. 4: 795–826.

Calhoun, Cheshire. 2000. "The Virtue of Civility." *Philosophy & Public Affairs* 29, no. 3: 251–275.

Collingridge, David. 1980. *The Social Control of Technology*. London: Pinter.

Colombetti, Giovanna, and Steve Torrance. 2009. "Emotion and Ethics: An Inter-(en)active Approach." *Phenomenology and the Cognitive Sciences* 8 (4): 505–526.

Epley, Kelly. M. 2015. "Care Ethics and Confucianism: Caring through *Li*." *Hypatia* 30, no. 4: 881–896.

Fan, Ruiping. 2010. *Reconstructionist Confucianism: Rethinking Morality after the West*. New York: Springer.

Frischmann, Brett, and Evan Selinger. 2018. *Re-Engineering Humanity*. Cambridge: Cambridge University Press.

Grosser, Benjamin. 2014. "What Do Metrics Want? How Quantification Prescribes Social Interaction on Facebook." *Computational Culture* 4. https://computationalculture.net/what-do-metrics-want/.

Hausman, Daniel M., and Brynn Welch. 2010. "Debate: To Nudge or Not to Nudge." *Journal of Political Philosophy* 18, no. 1: 123–136.

He, Chengzhou. 2018. New Confucianism, Science and the Future of the Environment. *European Review* 26, no. 2: 368–380.

Huang, Yong. 2015. "Confucianism and the Perfectionist Critique of the Liberal Neutrality: A Neglected Dimension." *Journal of Value Inquiry* 49, no. 1–2: 181–204.

Hutton, Eric. L. 2006. "Character, Situationism, and Early Confucian Thought." *Philosophical Studies* 127, no. 1: 37–58.

Hutton, Eric. L. 2014. *Xunzi: The Complete Text*. Princeton: Princeton University Press.

Ihde, Don. 1990. *Technology and the Lifeworld*. Bloomington: Indiana University Press.

Iltis, Ana S. 2012. "Ritual as the Creation of Social Reality." In *Ritual and the Moral Life*, edited by David Solomon, Ruiping Fan, and Ping-cheung Lo, 17–28. Dordrecht: Springer.

Kamm, Frances. 2015. *The Trolley Problem Mysteries*. New York: Oxford University Press.

Kamphof, Ike. 2017. "A Modest Art: Securing Privacy in Technologically Mediated Homecare." *Foundations of Science* 22, no. 2: 411–419.

Kim, Sungmoon. 2009. "Self-Transformation and Civil Society: Lockean vs. Confucian." *Dao* 8, no. 4: 383–401.

Kim, Sungmoon. 2010. "Beyond Liberal Civil Society: Confucian Familism and Relational Strangership." *Philosophy East and West* 60, no. 4: 476–498.

Kim, Sungmoon. 2012. "Before and After Ritual: Two Accounts of *Li* as Virtue in Early Confucianism." *Sophia* 51, no. 2: 195–210.

Kroes, Peter, and Anthonie Meijers, eds. 2000. *The Empirical Turn in the Philosophy of Technology*. Amsterdam: JAI-Elsevier.

Kroes, Peter, and Anthonie Meijers. 2016. "Toward an Axiological Turn in the Philosophy of Technology." In *Philosophy of Technology after the Empirical Turn*, edited by Maarten Franssen, Pieter E. Vermaas, Peter Kroes, and Anthonie Meijers, 11–30. Cham: Springer.

Kupperman, Joel. 2002. "Naturalness Revisited." In *Confucius and the Analects: New Essays*, edited by Bryan W. Van Norden, 39–52. New York: Oxford University Press.

Kupperman, Joel. 2004. "Tradition and Community in the Formation of Self." In *Confucian Ethics: A Comparative Study of Self, Autonomy and Community*, edited by Kwong-Loi Shun and David B. Wong, 103–123. Cambridge: Cambridge University Press.

Kupperman, Joel. 2007. "The Ethics of Style and Attitude." In *Moral Cultivation: Essays on the Development of Character and Virtue*, edited by Brad K. Wilburn, 13–28. Lanham, MD: Lexington Books.

Kupperman, Joel. 2010. "Confucian Civility." *Dao* 9, no. 1: 11–23.

Li, Chenyang. 2007. "*Li* as Cultural Grammar: On the Relation between *Li* and *Ren* in Confucius' Analects." *Philosophy East and West* 57, no. 3: 311–329.

McPherson, David. 2018. "Manners and the Moral Life." In *The Theory and Practice of Virtue Education*, edited by Tom Harrison and David Ian Walker, 140–152. New York: Routledge.

Meacham, Darian, and Matthew Studley. 2017. "Could a Robot Care? It's All in the Movement." In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, edited by Patrick Lin, Keith Abney, and Ryan Jenkins, 97–112. New York: Oxford University Press.

Mullis, Eric, C. 2007. "The Ethics of Confucian Artistry." *Journal of Aesthetics and Art Criticism* 65, no. 1: 99–107.

Neville, Robert Cummings. 2015. "Value and selfhood: Pragmatism, Confucianism, and Phenomenology." *Journal of Chinese Philosophy* 42, no. 1–2, 197–212.

Niedderer, Kristina. 2007. "Designing mindful interaction: the category of performative object." Design Issues 23, no. 1: 3–17.

Niedderer, Kristina. 2014. "Mediating Mindful Social Interactions through Design." In *The Wiley Blackwell Handbook of Mindfulness*, edited by Amanda Ie, Christelle T. Ngnoumen, and Ellen J. Langer, 345–366. New York: Wiley.

Nordmann, Alfred. 2007. "If and Then: A Critique of Speculative Nanoethics." *Nanoethics* 1, no. 1: 31–46.

Nordmann, Alfred, and Arie Rip. 2009. "Mind the Gap Revisited." *Nature Nanotechnology* 4: 273–274.

Olberding, Amy. 2015. "From Corpses to Courtesy: Xunzi's Defense of Etiquette." *Journal of Value* 49, no. 1–2: 145–159.

Olberding, Amy. 2016. "Etiquette: A Confucian Contribution to Moral Philosophy." *Ethics* 126, no. 2: 422–446.

Ott, Margus. 2017. "Confucius' Embodied Knowledge." *Asian Studies* 5, no. 2: 65–85.

Parviainen, Jaana, and Jari Pirhonen. 2017. "Vulnerable Bodies in Human-Robot Interactions: Embodiment as Ethical Issue in Robot Care for the Elderly." *Transformations* 29: 104–115.

Pols, Jeannette. 2017. "How to Make Your Relationship Work? Aesthetic Relations with Technology." *Foundations of Science* 22, no. 2: 421–424.

Puett, Michael. 2015. Ritual and Ritual Obligations: Perspectives on Normativity from Classical China. *Journal of Value Inquiry* 49, no. 4: 543–550.

Radice, Thomas. 2017. "*Li* (Ritual) in early Confucianism." *Philosophy Compass* 12, no. 10. https://doi.org/10.1111/phc3.12463.

Rosenberger, Robert, and Peter-Paul Verbeek, eds. 2015. *Postphenomenological Investigations: Essays on Human-Technology Relations*. London: Lexington Books.

Saito, Yuriko. 2017. *Aesthetics of the Familiar: Everyday Life and World-making*. Oxford: Oxford University Press.

Sarkissian, Hagop. 2017. "Situationism, Manipulation, and Objective Self-awareness." *Ethical Theory and Moral Practice* 20, no. 3: 489–503.

Seok, Bongrae. 2012. *Embodied Moral Psychology and Confucian Philosophy*. Lanham, MD: Lexington Books.

Sharon, Tamar. 2017. "Towards a Phenomenology of Technologically Mediated Moral Change: or, What Could Mark Zuckerberg Learn from Caregivers in the Southern Netherlands?" *Foundations of Science* 22, no. 2: 425–428.

Slaby, Jan. 2016. "Mind Invasion: Situated Affectivity and the Corporate Life Hack." *Frontier in Psychology* 7: 266. https://doi.org/10.3389/fpsyg.2016.00266.

Slingerland, Edward. 2003. *Confucius Analects: with Selections from Traditional Commentaries*. Indianapolis, IN: Hackett Pub Co.

Slingerland, Edward. 2011. “The Situationist Critique and Early Confucian Virtue Ethics.” *Ethics* 121, no. 2: 390–419.

Stalnaker, Aaron. 2016. “In Defense of Ritual Propriety.” *European Journal for Philosophy of Religion* 8, no. 1: 117–141.

Stohr, Karen. 2011. *On Manners*. New York: Routledge.

Sung, Winnie. 2012. “Ritual in Xunzi: A Change of Heart/Mind.” *Sophia* 51, no. 2: 211–226.

Thaler, Richard, and Cass Sunstein. 2008. *Nudge*. New Haven: Yale University Press.

Tronto, Joan C. 2005. “An Ethic of Care.” In *Feminist Theory: A Philosophical Anthology*, edited by Ann E. Cudd and Robin O. Andreasen, 251–256. Hoboken, NJ: Blackwell.

Tuuri, Kai, Jaana Parviainen, and Antti Pirhonen. 2017. “Who Controls Who? Embodied Control within Human-Technology Choreographies.” *Interacting with Computers* 29, no. 4: 494–511.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.

Van den Hoven, Jeroen. 2008. “Moral Methodology and Information Technology.” In *Handbook of Information and Computer Ethics*, edited by Kenneth E. Himma and Herman T. Tavani, 49–67. Hoboken, NJ: Wiley.

Verbeek, Peter-Paul. 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park: Pennsylvania State University Press.

Verbeek, Peter-Paul. 2011. *Moralizing Technologies: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press.

Wong, Pak-Hang. 2012. “*Dao*, Harmony and Personhood: Towards a Confucian Ethics of Technology.” *Philosophy & Technology* 25, no. 1: 67–86.

Wong, Pak-Hang. 2013. “Confucian Social Media: An Oxymoron?” *Dao* 12, no. 3: 283–296.

Yu, Erica, and Ruiping Fan. 2007. “A Confucian View of Personhood and Bioethics.” *Bioethical Inquiry* 4, no. 3: 171–179.

Zhao, Wenqing. 2014. “A Confucian Worldview and Family-Based Informed Consent: A Case of Concealing Illness from the Patient in China.” In *Family-oriented Informed Consent: East Asian & American Perspectives*, edited by Ruiping Fan, 231–244. Dordrecht: Springer.

CHAPTER 31

CARE ETHICS, PHILOSOPHY OF TECHNOLOGY, AND ROBOTS IN HUMANITARIAN ACTION

AIMEE VAN WYNSBERGHE

1. Introduction

Many of the emerging robots of our era are intensely social in nature. Designers of these robots intend not only to fulfil a service for humans but to create robots to enter into social relationships with humans. Just as social networking platforms promise to connect people by defying borders of time and space; social robots offer to fill a void of loneliness in people; sex robots promise to relieve the vulnerability associated with intimacy. These technologies are directed at social practices and confront us with significant "technomoral" choices (Vallor 2016, 2) concerning how to live well with technology when that same technology threatens to impoverish an individual's ability to achieve reciprocal relationships with other human beings. Reciprocity, to be sure, is the capacity for mutual care and simply put concerns an individual or society's ability to care for those who provide care. To account for this profound impact of technology on the ability to form reciprocal relationships, today's philosophy of technology must bring attention to human relationships from the periphery to the fore. Fortunately, the philosophy and ethics of care are equipped for this task (Gadow 1984, 2002; Held 2006; Noddings 2002; Reich 1995; Tronto 1993; van Wynsberghe 2013a; Vanlaere and Gastmans 2011; Verkerk 2001; Wilson 2002).

Care ethics is meant to provide an alternative vantage point to ethical and philosophical discourse, one that does not rely on consequences or absolute rules. Instead, care ethics stresses the centrality of three specific elements which are part of the process of uncovering and/or resolving (if possible) ethical issues: first, responsibility to one's self and to others; second, relationships between people; third, the many roles that people take on in their various relationships (Tronto 2013; 1993; Gilligan 1982). Each of these elements provides a starting point for ethical reasoning and when taken together provides a framework for shaping the interpretation of an ethical issue. These elements are bound together by something called a "care practice": the series of actions that take place in which caregiver and care receiver are brought together to provide, or receive, care respectively (Tronto 2013). It is within a care practice that a reciprocal care relationship is formed where both actors play an active role, that roles and responsibilities are (implicitly and explicitly) assigned, and that care actions are carried out (van Wynsberghe 2016a and 2016b). In essence, the significance of care practices (the space in which a relationship is formed) found in care ethics reformulates some[1] of the questions concerning technology to: "What is the impact of this technology on the development of a reciprocal relationship between caregiver and care receiver?"

Another central tenet of the care ethics tradition is an insistence that care happens on a more general and global level, in other words care exists outside the canonical examples of health care or child care (Tronto 2013). In this chapter we will visit an underexplored context in which both care and technology play a principal role: the humanitarian context. To be sure, this is not one context but a combination of morally charged scenarios ranging from refugee camps, detention centers, transition of migrants, and/or displaced persons during or following a natural disaster. What each of these share in common is the presence of vulnerable demographics in need of, and receiving, assistance from humanitarian organizations. This response to need, matching caring action with caring intention, is what we may call "care."

Technology is a central feature in the work of humanitarians in their relief efforts. Robotics is one of the innovations the humanitarian sector is looking to pursue to overcome the obstacles facing humanitarian workers. Robots in the humanitarian space (the most pervasive being drones [van Wynsberghe and Comes 2019]) have been met with both criticism and praise, largely by practitioners and UN organizations but also scholars and activists (Soesilo et al. 2016; Raymond et al. 2012; Choi-Fitzpatrick 2014; Clarke 2014; Karlsrud and Rosén 2013; UN-OCHA 2014; van Wynsberghe and Comes 2019; Chow 2012; Sandvik and Jumbert 2016; van Wynsberghe et al. 2018). On the one hand, many believe the technology will exacerbate existing organizational and logistical issues. On the other hand, many believe robots can solve problems that humans can't, including last mile deliveries to locations that humans can't physically reach or the provision of aid during viral outbreaks such as the Ebola or Corona virus crises, when human contact is dangerous.

This chapter uses the philosophy of care to enhance the philosophy of technology in addressing the nature of technology in the humanitarian space. A result of this will be better evaluations of technologies within humanitarian contexts by focusing on the impact the technology has on the development of reciprocal relationships between caregiver (humanitarian actor) and care receiver (i.e., beneficiary, refugee, detainee, patient, etc.).

The chapter begins with a discussion of the role the philosophy of care can play within the philosophy of technology, an exercise which in turn elaborates the central features of care ethics previously mentioned. This will provide the humanitarian discourse with a set of conceptual tools to describe, construct, and evaluate the practices constitutive of humanitarian action. The chapter continues with an exploration of the current robot prototypes being tested and proposed for humanitarian action. After demonstrating the strength of care ethics in highlighting that certain technomoral choices for humanitarian actors exist and should be accounted for, the chapter continues with two thought experiments[2] to encourage moral learning on what care ethics would ask of us in specific humanitarian contexts with explicit robot applications: first, the detention center context in which a drone is used to inspect conditions from the air; second, an Ebola crisis in which a tele-operated robot is used to deliver resources to ill care receivers in isolation.

The chapter presents three main points to the reader that deserve attention. First, the concept of care deserves further treatment within the philosophy of technology insofar as technologies are nested within care practices; second, in the same way that the home, the school, the hospital, or nursing home have been discussed as a place for care, the humanitarian space must also be discussed as a place of care and subject to evaluations that are grounded in the care ethics tradition; and third, robots in humanitarian action should be evaluated according to their ability to strengthen or weaken the possibility for reciprocal human relationships.

2. Philosophy of Technology with Care

One of the 20th century's most cited philosophers of technology, Martin Heidegger, was inspired by the concept of care and its relation to "being-in-the-world" (Heidegger 1996); however, current discussions within the philosophy and ethics of technology rarely return to this idea. Recent discussions by philosopher of technology Shannon Vallor focus on virtue ethics to deliberate how technology can be used to assist in human flourishing, character and/or excellence development (Vallor 2016).[3] Alternatively, philosopher of technology Albert Borgmann engages in an assessment of technology by making a plea for directing our technologies toward a life of engagement, the foundation for which is built from practices to cultivate personal excellence (Borgmann 1987). Both of these approaches privilege personal accomplishments of individualistic human beings rather than a recognition of the primacy of relational human beings; as such, they neglect accomplishments in terms of reciprocity in relationships. Care ethics, in focusing on relationships, tethers the central ethical question of introducing technology to the dominance of relationships and the value of reciprocity within these.

Many contemporary philosophers of technology have argued that there are often (or always) prescriptive factors at work during the creation, development and deployment of technologies. For instance, philosophers have elaborated on the prescriptive role of engineers and/or designers and have suggested that technologists "materialize morality" insofar as their own assumptions, biases, and values find their way into the architecture, capabilities, and even materials used to create technologies (Verbeek 2006; 2011). More often than not, however, this dialogue neglects the assumptions, biases, and norms associated with "care" and the consequential embedding of these variables into technologies. Omission of this important consideration has gained attention in the nursing sciences (Barnard and Sandelowski 2001; Wilson 2002; Watson 1999; 2011; Gadow 2002; 1984; Marian Verkerk 2001); however, these discussions are restricted to health care contexts, health care technologies, and often to (mostly female) nurses. If we tend to think of care as a female specialty in contrast with technology (development or use), it is because we have been encouraged to do so.

To counter this trend, scholars of the philosophy of care have worked hard to demonstrate the prevalence of care beyond the hospital. They do this by illustrating the variety of care practices that occur in the lives of individuals on a daily basis. These care practices are defined neither by place alone (e.g., health care, home, or school) nor by person alone (i.e., parent, nurse, teacher, female). In other words, care is much broader than the hands-on work by a few select individuals in specific contexts. Of equal importance, these scholars show the role that technology plays in the provision of care (Tronto 2013; 2010; 1993; Held 2006; Verkerk 2001; Noddings 2002; van Wynsberghe 2016a and 2016b; van Wynsberghe and Gastmans 2008). Most of these authors address situations in which technologies are already in place and/or is required to reach the ends of care. As such, technology is a part of caring. With this in mind, let us take a closer look at the philosophy of care to assess its applicability for evaluating the role of technology within care.

2.1 Care Revisited

The roots of the care ethics tradition can be traced back to Carol Gilligan and her book *In a Different Voice* (Gilligan 1982). This book, and the reflections it provides, essentially paved the way for an alternative framework to moral reasoning (and moral development models) than that embodied in the Kantian or utilitarian frameworks that dominated the academic landscape of the time.

Building on the work of Gilligan, Berenice Fisher and Joan Tronto argue that good care requires both action and intention. Moreover, that care should be defined as "a species activity that includes everything we do to maintain, continue, and repair our world so that we can live in it as well as possible" (Fisher and Tronto 1990, 40; see also Tronto 1993, 103). The strength of this definition is in showing the ubiquity of care; essentially care is part of everything we do. This is not to say that we care well in every instance; rather, care is very much a part of who we are and what we do on a daily basis.

Central to Tronto's conception of care is the understanding that good care begins with the needs of another, is relational, and is manifest through actions and intentions. These actions are not isolated events or tasks but are complex practices in which numerous actors are involved in the identification of needs, the meeting of needs, and the response to needs. These complex processes known as care practices have four iterative steps:

"1. *Caring about;* someone or some group notices unmet caring needs.
2. *Caring for;* someone or some group has to take responsibility to make certain that these needs are met.
3. *Caring-giving;* actual care-giving work is done.
4. *Care-receiving;* throughout the provision of care work the response from the person, thing, group, animal, plant, or environment is observed and together judgments are made about it" (Tronto 1993, 22–23).

This last phase of care-receiving highlights the importance of responsiveness, that care should not happen *to* someone (or something); rather, care happens *with* someone (or something). Care receivers have a role to play in the care process—they have opinions, preferences, desires, and hopes; they are holistic persons with life histories that should be shared as part of good care (Vanlaere and Gastmans 2011). Recognizing this is a step toward re-calibrating the asymmetry in power between the powerful caregiver and the "vulnerable" care receiver (Verkerk 2001).

Responsiveness in turn highlights another central element in care, namely reciprocity. Both responsiveness and reciprocity are features of the active role of care receivers, reminding us that care is bi-directional rather than uni-directional. Responsiveness and reciprocity are, however, not the same. Responsiveness implies a reaction to the care being provided, indicating "yes this new position is more comfortable" or, "please remove the bedpan." Reciprocity is about giving back to caregivers for the care they have provided; about mutually caring for those who provide care. Mutual care can be achieved through a variety of gestures, for example saying thank you or returning a favor. It is essentially a sign of appreciation for the care provided. Moreover, reciprocity is more than just a sign of appreciation in the short term or on a small scale, it is about mutually caring for caregivers across society and for the long-term. It is about ensuring that caregivers are not financially (or otherwise) weakened merely because of their role as caregiver.

To be sure, reciprocity is not discussed in the care ethics tradition alone but has a rich history in a variety of disciplines, e.g., moral development (Duska and Whelan 1975) and sociology (Gouldner 1960) to name a few. However, both responsiveness and reciprocity share a privileged position in the care ethics tradition; both concepts highlight the relationality of humans. It is from such a relational lens that one ought to, according to care ethicists, interpret other ethical values and principles such as autonomy of persons. It is care ethicists who reject traditional atomistic conceptions of autonomy and suggest instead that autonomous decision making, of patients for example, can only be understood as a balancing of roles and responsibilities that an individual has in relation to others in his/her life (Verkerk 2001).

For care ethics, reciprocity is the foundation for the valuation of caregivers and the work that they do. It is a recognition of both the relational and bi-directional nature of care, and it is integral to the just society. It follows then that evaluating technology according to its ability to manifest, or at the very least protect, the value of reciprocity must take center stage for a philosophy of technology that pays tribute to the tenets of care ethics.

2.2 A Caring Philosophy of Technology

In Tronto's most recent book *Caring Democracies: Markets, Equality, and Justice* (2013) she sets out to explore the revolution in care observed over the last century. Over the last 100 years care as a concept has shifted from one rooted in the home setting [i.e., care happens between parents, children, and siblings, as found in the oft cited works of (Noddings 2002)] to one that extends beyond the private space of the household. Over time, caring institutions such as schools, hospitals, hospices, nursing homes, and homes for the disabled have been created, culminating in the professionalization of care. Tronto calls this new form of care, democratic care, and argues that this public form of care requires a commitment to the relational aspect of care in so far as good care cannot be decided from one vantage point alone[4] (Tronto 2013). In Tronto's articulation of democratic care—that care exists in the public sphere—she adds a fifth phase of care to the four listed earlier: "5. *Caring with;* care requires that caring needs, and the ways in which they are met, must be consistent with democratic commitments to justice, equality, and freedom for all" (Tronto 2013, 23). A commitment to mutual care for caregivers is the kind of justice that Tronto puts forward—care as a political ideal recognizes and promotes reciprocity to those who provide care so as to ensure that caregivers are not weakened in society by the reality that they are caregivers.

This movement of care from private to public is important for the philosophy of technology as it opens up a plethora of technologies that fall within the scope of care ethics. To be sure, while the place of care is not restricted to private settings alone (it has become public and professionalized), there have long been public and professionalized forms of care in many cultures, from battlefield care to care of animals, to care of the land, to care for the rule of law (i.e., "duty of care"). My point here is not necessarily that the scope of care has widened *in fact*; rather, we as philosophers are now able to widen (as a correction) what we can now *recognize* to be the true scope of care. This recognition of a broader scope, this more public form, of care demands strengthened commitments to democratic values of justice, equality, and freedom for all. What's more, this widening of the scope of care broadens the possible kinds of technologies to be assessed; the omnipresence of care means that the technologies that enter our lives are increasingly entering care practices. As such, care ethics is an essential, foundational and central element of a proper philosophy of technology. The question we might ask now is what, if any, caring roles will be ascribed to these technologies and how reciprocity within the caring relationship is preserved, enhanced, or diminished with the addition of a certain technology.

3. Robots in Humanitarian Action

The push to increase the use of emerging technologies in the humanitarian space was recently strengthened at the World Humanitarian Summit: “the 2016 Agenda for Humanity of the United Nations Secretary General states that to deliver collective outcomes, the humanitarian sector must promote a strong focus on innovation” (Soesilo et al. 2016, 10). Robotics is one such innovation high on many humanitarian organizations’ list of priorities.

For the reader who considers such a discussion premature it is true that prototypes and applications of robots in the humanitarian space are still somewhat minimal; however, there is reason to believe this won’t be the case for long. Attention and pressure began to build during the 2014 Ebola outbreak in West Africa during which work to pursue various kinds of robots for humanitarian efforts intensified. On November 7, 2014 scientists met at universities across the United States with the help of the White House Office of Science and Technology Policy to discuss the role that autonomous machines might play in combating Ebola crises (Ackerman 2014; Copestake 2014). Discussions led to a variety of robot prototypes (some already commercially available and others not) which were tested for use in humanitarian contexts such as: those for killing germs and pathogens like the Ebola virus (e.g., Xenex Germ-zapping robot that destroys Ebola and Anthrax spores); robots for removing people from battlefield or conflict zones (e.g., the US Army’s Black Night Transformer); robots for performing long distance, remote robotic surgery on people in the battlefield or conflict zones (DARPA Trauma pod; the Zeus Telesurgical system or the daVinci robotic system); and robots for removing debris.

There are other ideas for robots in the humanitarian space including, but not limited to, robots for: delivery of food and medicine to the sick; delivery of organs, vaccines, and blood samples for medical treatment; decontamination of equipment; and the burial of the dead. In 2013, DARPA hosted the Robotics Challenge and created a series of tasks based on the 2011 Fukushima nuclear disaster so that robots could be programmed to open doors, drill holes in walls, manipulate valves and other similar tasks.

One of the most prominent robots to be found in the humanitarian space is the unmanned aerial vehicle, also known as the drone. The first drones deployed in this sector were used for surveillance in peacekeeping missions in the Democratic Republic of Congo in 2006 (Karlsrud and Rosén 2013) and in 2013, the first cargo drones were used to deliver medical supplies (Choi-Fitzpatrick 2014). With the growth in numbers and applications,[5] in 2018 UNICEF hired its first dedicated specialist to coordinate drone projects and testing. According to the FSD report on drones, the most promising uses of drones in the humanitarian sector include: “mapping, delivery of essential items to remote or hard-to-access locations, supporting damage assessments, increasing situational awareness, and monitoring changes” (Soesilo et al. 2016, 7). Drones also show great potential in tactical settings as support for the work of search and rescue teams and field teams” (Soesilo et al. 2016, 19).

In short, there has been a considerable increase in the number of robots used, tested, and proposed in the last 10 years for this sector and we have reason to believe this trend will continue in the years to come. There are a variety of reasons to proceed with the use of robots in humanitarian contexts with caution: the experimental nature of including any new technology into a social context (van de Poel 2013; 2018); the vulnerability of the demographics who will be impacted by robots in the humanitarian space; and the risk that robots will be perceived as the "solution" to the breadth of problems facing the humanitarian sector. In short, the stakes are high and any responsible implementation of robots in the humanitarian sector should be preceded by a critical evaluation to ensure its implementation has clear goals and is able to achieve said goals. In what follows, we will explore the role that care ethics plays in strengthening the humanitarian principles approach.

4. Enriching Humanitarian Ethics using Care Ethics

Humanitarian action, as defined by Hugo Slim (who is both a humanitarian scholar and practitioner), is considered "a compassionate response to extreme and particular forms of suffering arising from organized human violence and natural disaster" (Slim 2015, 1). It is about "protecting, respecting, and saving human life" (Slim 2015, 2). Merging this definition with that of 'care,' provided earlier, might suggest the following:

> Humanitarian action concerns the variety of care activities (e.g. communication, delivery of goods, medical care, personal development, environmental resilience building, and more) performed to repair the world in a specific space, the humanitarian space.

It is care directed at those in life-threatening need, and done with a compassionate disposition; it is action aligned with disposition. Consequently, humanitarian action may rightly be considered as a series of care practices.

In light of serious ethical dilemmas and conflicts that humanitarian workers face at the hands-on and institutional levels, the Fundamental Principles of the Red Cross and Crescent Society were agreed on in 1965. These are commonly recognized by the humanitarian sector as the 'humanitarian principles.' They represent the core of humanitarian ethics and are as follows: "*humanity*, human suffering must be addressed wherever it is found; *impartiality*, humanitarian action must be carried out on the basis of need alone; *neutrality*, humanitarian actors must not take sides; and *independence*, humanitarian action must be autonomous from political, economic, or military objectives" (Bagshaw 2012).

While the humanitarian principles provide overarching guidance on the aim of humanitarian care in general, they require another lens to provide guidance on the evaluation of specific technologies integrated into humanitarian care practices. For instance, the principle of humanity is clear insofar as humanitarian action must work to alleviate human suffering, but this says little about how to systematically evaluate a technology for its ability to do so. It is for this gap that the ethics and philosophy of technology can provide assistance as a bridge; they offer a lens for evaluating the broader meaning and/or impact of technology. More specifically, this is precisely where the care ethics tradition can provide a rubric for evaluating a technology according to its ability to promote or threaten the principle of humanity. It can do this by using the notion of reciprocity in the relationship formed between caregiver and care receiver.

In keeping with the empirical turn within the philosophy of technology (Achterhuis 2001), the following two scenarios will explore concrete robot applications in the humanitarian space. The scenarios will guide the reader through separate care practices in the humanitarian space that may involve robot assistance or substitution in the near future.[6] The two plausible instances are based on uses for which robot prototypes are currently being explored or those for which there is a high likelihood for robots to be applied in the near future (based on applications outside the humanitarian sector). Each of these scenarios will be assessed from the care ethics perspective to explore: the impact of the prospective robot on the establishment or maintenance of a reciprocal relationship between caregiver and care receiver; to what extent the robot can be beneficial to the goals of the humanitarian care practice; and to what extent the robot might be detrimental to the goals of the humanitarian care practice.[7] Lastly, the scenarios will show more concretely what indispensable reflections care ethics adds to the toolkit of the philosophy of technology.

4.1 Drones to Inspect Detention Centers

Drones were first introduced to the humanitarian sector in the early 2000s as a surveillance technology,[8] and they have since become far more accessible, with off-the-shelf products available for around US $200 or less (Soesilo et al. 2016). Alongside the pervasive use of drones in the humanitarian space, the drone has become a target for private industry. In a recent report conducted by the global accounting firm PricewaterhouseCoopers (PwC), the use of drones for inspection is anticipated to reach a target of $45.2 billion globally (Mazur and Wisniewski 2016). Given the factors that combine to make humanitarian drones attractive—a shortage of humanitarian personnel, an increase in affordable off-the-shelf drone technology, and an increase in drone applications for building maintenance—let us imagine a future use case of a drone for the inspection of detention centers in the place of human humanitarian actors.

There are a variety of places of detention that the International Association of the Red Cross (ICRC) is committed to investigating: "In 2012, ICRC delegates visited about 540,000 detainees in 97 countries and territories, more than 26,000 of whom were seen in private interviews" (Bouvier 2012). Reasons for holding a person in a detention center may be to serve a sentence after breaking the law, or immigration detention (when a person does not have the right papers to enter a country or is living in a country illegally). No matter the reason for their arrest or detention, "the ICRC aims to secure humane treatment and conditions of detention for all detainees."[9] In these instances the ICRC have workers who "work as an impartial, independent, and neutral organization within the framework of private, confidential interviews with detainees, and of a confidential dialogue with the detaining authorities" (Bouvier 2012, 1258). Aid workers visit detention centers to assess living conditions, take reports of abuse and torture of detainees, and in the process try to advocate for changes in these conditions while also providing moments of positive experiences to the detainees.

In a 2012 opinion piece published by the ICRC, medical doctor Paul Bouvier (who has participated in many such detention center visits) describes the value of these visits for raising spirits of detainees, reaffirming their dignity, or just sharing a moment of humanity. These moments can be as simple as sharing a cup of coffee, tea, or a cookie with a detainee or more engaged human activities such as sharing pictures, news from home, or other small gifts. Consequently, these hands-on practices serve multiple (institutional) ends for the humanitarian organization; they assist the aid worker in understanding and interpreting the situation (e.g., what kind of space is being provided, what kind of support or lack thereof is being provided and so on) in addition to providing respite for the detainee in often inhumane living conditions.

In his piece, Bouvier confronts the feeling of powerlessness experienced in conducting a detention visit, asking if it is enough to share a cup of coffee during a short visit when the detainee lives in such horrible conditions. Or worse, should these visits continue if they do not bring an end to the torture of the detainees?[10] One might guess that such a line of questioning on the part of Bouvier may arise from a worry that such visits implicitly legitimize or unwittingly perpetuate the inhumane circumstances. And it is along those lines that humanitarian organizations could find situations for which using a drone may be justified: the remote inspection of the physical structure of the detention center, visual representation of the living conditions of detainees that could be captured through images taken by the drone, and an escape from the feeling of futility of such visits. In essence it could be asserted, on behalf of humanitarian organizations, that use of a drone would be a more efficient use of resources to achieve the goal of inspecting detention centers. And in violent situations, it could be argued that substituting a drone for a humanitarian worker would protect the safety of the worker.

When unwrapping this care practice we see that the care worker (the ICRC worker) must assess living conditions, make reports, and engage in dialogue with stakeholders other than the care receiver (i.e., those in command of the detention center). But the practice also serves as a vehicle to establish a relationship between caregiver (ICRC visitor) and care receiver (detainee) aimed at respecting the dignity of the detainees (i.e., realizing the principle of humanity). It is in the process of establishing the relationship, through dialogue and shared moments, that the dignity of the care receiver is realized.

Of equal importance, in these engagements the care receiver (detainee) often reciprocates their appreciation to the visitor by sharing stories about their life. This requires trust on the part of the care receiver in a time when their living situation does not drive them to trust easily—already an act of reciprocity. In some instances detainees expressed greater acts of reciprocity—gestures of mutual care like offering drink, food, a drawing or a poem—not out of "moral obligation ... but rather as an expression of gratitude and a request for recognition as a human being with an identity and a history, emotions, sufferings and capacities" (Bouvier 2012, 1542). Thus, reciprocating to caregivers not only confirms the value of the work of the caregiver but also affirms the dignity of care receivers and their ability to provide care for others (as opposed to being merely recipients of care).

The technology therefore cannot be evaluated according to its impact on either caregiver or care receiver, instead it must be evaluated according to its impact on the ability to establish a reciprocal relationship throughout the practice of detention center visits. Upon understanding the ends that the practice serves (e.g., assessment of living conditions, treatment of detainees, and confirmation of the dignity of detainees through the ability to act reciprocally), the technomoral choice depends on the answer to the following question: Will the use of a drone threaten the development of a reciprocal relationship between the humanitarian caregiver and the detainee (care receiver)?

From the preceding evaluation we can see that using the drone to replace a human visitor to the detention center will threaten the establishment of a reciprocal relationship between detainee and visitor. In other words, the importance of maintaining the reciprocal, caring relationship between detainee and visitor demands an alternative, more humane and ethical use of a drone for inspection of detention centers. If the drone is to be used at all, perhaps it may be introduced into the overall practice of detention center visitation as an assistant to the visitor, as a means to provide security to the visitor in a risky situation. The role and/or responsibility delegated to the drone is that of inspection and/or visual documentation of living facilities alone. In such a scenario the drone is able to capture images of the detention center for the record while the caregiver is inside engaging in meaningful interactions with the detainees.

In short, this scenario illustrates that the care ethics perspective can provide a way of understanding detention center visitation as a care practice. In so doing it is possible to identify how and why values such as reciprocity (or the humanitarian principle of humanity) come to be expressed. With this in mind it is then possible to evaluate the introduction of the robot beyond how a robot will interact with a human user; rather, it is possible to identify what will happen if/when a robot impedes humans from meaningful, reciprocal, interaction with one another.

4.2 Robot Nurses during an Ebola Crisis

In November 2014, experts from the United States met to discuss mitigation strategies for dealing with the Ebola crisis in West Africa at the time. Two epidemiologists, Michelle Dynes and Anne Purfield, were among them, having recently returned to the United States from Sierra Leone, where they had served as aid workers during the Ebola crisis. They shared a story of twelve nurses who were infected with, and died from, the disease after providing physical comfort to a baby who had been separated from her mother because the mother was also infected with the Ebola virus. The disease spreads through human contact, making it vital to separate the uninfected from carriers and to sterilize, then burn, the infected bodies of the dead. In an article discussing the technological possibilities to manage the disease, even the title suggests a solution, "The Best Nurses for Ebola: Robots?" The article goes on to suggest, "From a technological standpoint, the best way to combat all of this is for the healthy to distance themselves from the stricken. And the most obvious way to do that is to remove human interaction from the equation. And the most obvious way to do *that* may involve removing humans themselves from the equation—at least when it comes to the care of the sick" (Garber 2014). Several technologies could achieve this goal, including telepresence robots designed to replace suit-wearing human workers or autonomous delivery robots deployed to bring food and medicine around a quarantine care facility, to name two. For the sake of this scenario, let us imagine a sterile delivery robot that travels throughout a care facility to deliver sheets, medications, food, and other items to patients in sterile environments.

One can envision a room in which one or more infected patients (ideally one but if space is not permitting, then multiple) remain for the duration of their infection. At multiple times of the day humans must enter the rooms to bring food, medicine, and other supplies. During these visits they will engage in conversation with the care receivers too. In situations of viral outbreak, such as an Ebola crisis, caregivers will wear a protective suit from head to toe as a preventive precaution. Given the risk of contracting this fatal virus and rather than having a human in a protective suit enter with supplies (food, medicine or others), a delivery robot could be used to bring materials throughout the facility, possibly saving the lives of caregivers.

Unpacking the practice of item delivery to the rooms of patients reveals more than just the distribution of resources, for there is also often physical interaction and verbal communication between caregiver and care receiver. In the same way that other daily practices, such as lifting, bathing, and feeding of patients, act as moments for establishing a relationship between caregiver and care receiver (van Wynsberghe 2013a; 2016a), the same can be said for the moments in which meals, medications, and/or sheets are delivered to a patient's room. It is in these moments that caregivers assess the medical status of the patient and can find opportunities for care receivers to respond to the care provided (i.e., to indicate whether their needs are being met, they need to be repositioned, or have the curtain closed, etc.).

These moments also serve as a mechanism for fostering reciprocity between caregiver and care receiver, whereby care receivers can show their appreciation for the care provided by engaging in meaningful discussions in which they express their wishes, hopes, fears, desires, and so on. Engaging in these conversations is a sign of care towards the caregiver, since we do not engage in meaningful dialogue with those we do not respect or appreciate. In other words, these moments for dialogue and relationship building should not be dismissed as trivial or expendable moments; rather, these are the moments in which care receivers have the opportunity to show appreciation toward their caregivers through relationship-building conversations. Moreover, as we saw from the earlier scenario, detainees willingly provided gestures of reciprocity to visitors, not out of moral obligation but out of a wish to be seen as more than a detainee. Patients that must be kept isolated may also wish to be seen as more than a patient, and one way to do this is to ensure that they can engage in conversation beyond merely sharing an update on their symptoms. Providing the opportunity for such moments (of appreciation from care receiver to caregiver) is one way in which this can happen.

If we recall, at the beginning of this scenario the aim was to explore the use of robots as a means to protect the lives of health care workers in a pandemic response. We should assume that the need for reciprocity does not demand that caregivers risk their lives, thus robots should be used if the wellbeing of nurses is at risk. However, using a robot for delivery will mean that care receivers lose opportunities for interactions with a human caregiver as human-human contact will be replaced with human-robot interactions. In such an instance there is a threat of impersonal care provided by a robot (care actions detached from caring dispositions) but also a loss of opportunities for reciprocity that are mutually beneficial for caregiver and care receiver. But perhaps the use of robots in a pandemic response need not be so restrictive. Perhaps there is a middle ground using care ethics, and the analysis provided above, to guide the capabilities of the robot and/or its mode of implementation into the health care system.

If designers and/or implementers (e.g., hospital managers) were to take seriously the need to preserve human interaction, and more specifically reciprocity between human care receivers and human caregivers, the robot's functioning may differ. The robot could be tele-operated by a care worker, family, or friend, meaning the robot's entrance into the hospital room would be controlled by one of the above individuals and the individual operating the robot could be stationed at a different area of the hospital, in a different building, or even on a different continent, effectively keeping the two humans safely separated. This first step would serve the purpose of maintaining safety of human caregivers but an additional step is needed to introduce a relational element to the practice. For this, it is possible to make the operator of the robot visible to the care receivers with the addition of verbal communication. In this way, the caregiver operating the robot at a distance can be seen by the care receiver in his/her hospital room and the two can engage in conversation while the robot fulfils its task (e.g., vacuuming the floor, sterilizing the room, emptying the waste bin etc.). Providing visibility of this kind has already been shown in preliminary studies to provide comfort in Ebola crisis situations (Kraft and Smart 2016). Adding the ability for care receivers to speak directly to a caregiver opens up the possibility for care receivers to reciprocate appreciation to their caregivers in whatever way possible.

In short, this scenario shows an instance in which care receivers could be forced into minimal human interactions (thereby threatening the promotion of humanity) as a security measure to protect caregivers.[11] If, however, in such instances care is taken as a starting point to uncover novel ethical concerns (e.g., concern for diminished reciprocity) and at the same time as motivation to mitigate such concerns, the introduction of the robot may be designed and/or deployed to uphold care values rather than threaten them. In this scenario, a robot used in pandemic response to deliver items to or sterilize a patient's room, could be explicitly designed and used to maintain (or introduce) opportunities for social interaction through technical capabilities, e.g., tele-operation, visibility of the tele-operator, and ability to engage in conversation between caregiver and care receiver. The idea of using ethics to steer innovation is certainly not new; however, the idea to prioritize reciprocal caring relationships on par with safety and/or security is new. Given the ubiquity of care practices across society and the prevalence of technology entering into these practices, it is time now for care ethics, and the values central to care ethics, to take a privileged role within the philosophy and ethics of technology.

5. Conclusion

In this chapter, the notion of care as a resource for the philosophy of technology and as a source for ethical evaluations was shown using scenarios of prospective robot applications in humanitarian action. Using the lens of care ethics to illustrate the care practice in its entirety, it was possible to conclude that the technology cannot be evaluated according to its impact on either caregiver or care receiver, instead it must be evaluated according to its impact on the ability to establish a reciprocal relationship throughout the practice.

In each of the scenarios presented here, using robots to keep isolated people connected with others seems obviously good to do, so why wouldn't we do that anyway? To be sure, there are mounting commercial and institutional pressures to under-invest in these kinds of humane design features, and to prioritize things like resource delivery, physical security, and surveillance. In the absence of a moral priority given to the caring relationship, where that is the *first* and *essential* item to preserve, then the likelihood is that humanitarian robots will be built with narrower values in mind: keeping people alive and uninjured, as cheaply as possible. Or monitoring violations of humanitarian law, as efficiently and cheaply as possible. To build in the humane features I have described here costs time and money, and requires other kinds of operational support. It's far easier and cheaper to design a robot that just sterilizes the room than one that also enables satisfying interpersonal contact. Care ethics shows that we have a greater obligation than we realize to prioritize the features of care ethics in the design and implementation of humanitarian technology, which existing utilitarian or principle-based frameworks would probably allow to be sacrificed to existing institutional and economic pressures.

Acknowledgments

This research is supported by the Netherlands Organization for Scientific Research (NWO), project number 275-20-054.

Notes

1. I say "some" questions of technology to assure the reader I am not asserting that all ethical questions of technology ought to be addressed in this manner. Rather, I am making a more modest claim and suggesting that technologies that may directly impact on the ability to form and/or sustain human relationships could benefit from the care ethics lens.
2. Thought experiments are one of three types of moral experiments described in Van De Poel (2018) with the aim of stimulating moral learning about new technologies. For a more elaborate description and critique of moral experiments, please see Van De Poel (2018).
3. It should be noted that a relationship between the virtue ethics approach and care ethics has been discussed in the work of Vallor, who states that care is a special kind of virtue, or a cluster of virtues, rather than a subject in need of separate ethical treatment. The starting point of this chapter rests on the belief that care ethics differs from virtue ethics insofar as the focus of the former is on the dyadic relationship between caregiver and care-receiver rather than on the skills of one or the other. Care ethics insists that proper ethical attention must be directed at care practices and the components within, rather than on one or the other human actors.
4. Another important aspect of democratic care for Tronto is understanding the centrality of the needs of every citizen (rather than those in need of care at one particular moment).
5. Although the humanitarian sector has experienced a surge in the use of drones in parallel to the consumer industry, an estimate of the number of drones in operation today cannot be found in current literature. This may be so for a variety of reasons, among which is the absence of requirements to register the use of drones with an overarching regulating body.
6. For similar evaluations of robots in health care contexts see (van Wynsberghe 2013a; 2013b).
7. To be sure, this is not an exhaustive list of all the ethical issues that will arise from these robot prototypes, there are issues related to data collection and usage, privacy, information transparency, physiological and behavioral impact of drones, experimentation on vulnerable demographics, and more.
8. The first drones deployed in this sector were used for surveillance in peace-keeping missions in the Democratic Republic of Congo in 2006 (Karlsrud and Rosén 2013).
9. For more on the work of ICRC, visit https://www.icrc.org/en/what-we-do/visiting-detainees (retrieved Feb. 4, 2019).
10. If the visits of the ICRC do not result in a change of conditions for detainees, the ICRC can decide to stop the visits altogether and publicly denounce the situation.
11. One could also imagine the detention center as a similar situation; however, the detainees are intentionally kept from their family and friends, and the medical example differs in this respect.

References

Achterhuis, Hans. 2001. *American Philosophy of Technology: The Empirical Turn*. Bloomington: Indiana University Press.

Ackerman, Evan. 2014. "Real Robots to Help Fight Ebola: IEEE Spectrum." *IEEE Spectrum: Technology, Engineering, and Science News*, October 20, 2014. https://spectrum.ieee.org/automaton/robotics/medical-robots/real-robots-to-help-fight-ebola.

Bagshaw, Simon. 2012. "OCHA on Message: Humanitarian Principles." http://www.unocha.org/sites/dms/Documents/OOM-humanitarianprinciples_eng_June12.pdf.

Barnard, Alan, and Margarete Sandelowski. 2001. "Technology and Humane Nursing Care: (Ir)Reconcilable or Invented Difference?" *Journal of Advanced Nursing* 34 (3): 367–375. https://doi.org/10.1046/j.1365-2648.2001.01768.x.

Borgmann, Albert. 1987. *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago: University of Chicago Press.

Bouvier, Paul. 2012. "Humanitarian Care and Small Things in Dehumanised Places." *International Committee of the Red Cross* (December). https://www.icrc.org/en/international-review/article/humanitarian-care-and-small-things-dehumanised-places.

Choi-Fitzpatrick, Austin. 2014. "Drones for Good: Technological Innovations, Social Movements, and the State." *Journal of International Affairs* 68 (1). https://search.proquest.com/openview/468bc87b04291e1f45ff0f60f9edf97b/1?pq-origsite=gscholar&cbl=41938.

Chow, Jack. 2012. "The Case for Humanitarian Drones." *OpenCanada*, December 12, 2012. https://www.opencanada.org/features/the-case-for-humanitarian-drones/.

Clarke, Roger. 2014. "The Regulation of Civilian Drones' Impacts on Behavioural Privacy." *Computer Law & Security Review* 30 (3): 286–305. https://doi.org/10.1016/j.clsr.2014.03.005.

Copestake, Jen. 2014. "Ebola Robots on White House Agenda." *BBC News*, November 7, 2014, sec. Technology. https://www.bbc.com/news/technology-29942392.

Duska, Ronald, and Mariellen Whelan. 1975. *Moral Development: A Guide to Piaget and Kohlberg*. Dublin: Gill and Macmillan.

Fisher, Berenice, and Joan Tronto. 1990. "Toward a Feminist Theory of Caring." In *Circles of Care: Work and Identity in Women's Lives*, edited by E. Abel and M. Nelson, 35–62. Albany, NY: SUNY Press.

Gadow, Sally A. 1984. "Touch and Technology: Two Paradigms of Patient Care." *Journal of Religion and Health* 23 (1): 63–69.

Gadow, Sally A. 2002. "Nurse and Patient: The Caring Relationship." In *Caring, Curing, Coping: Nurse, Physician, and Patient Relationships*, edited by A. H. Bishop and J. R. Scudder, 31–43. Tuscaloosa, AL: University of Alabama Press. http://books.google.nl/books?id=MOWjAr6lB2oC.

Garber, Megan. 2014. "The Best Nurses for Ebola: Robots?" *The Atlantic*. October 24, 2014. https://www.theatlantic.com/health/archive/2014/10/the-best-nurses-for-ebola-patients-might-be-robots/381884/.

Gilligan, Carol. 1982. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.

Gouldner, Alvin W. 1960. "The Norm of Reciprocity: A Preliminary Statement." *American Sociological Review* 25 (2): 161–178. https://doi.org/10.2307/2092623.

Heidegger, Martin. 1996. *Being and Time: A Translation of Sein und Zeit*. Albany, New York: SUNY Press.

Held, Virginia. 2006. *The Ethics of Care: Personal, Political, and Global*. New York: Oxford University Press.

Karlsrud, John, and Frederik Rosén. 2013. "In the Eye of the Beholder? UN and the Use of Drones to Protect Civilians." *Stability: International Journal of Security and Development* 2 (2): 1–10. https://doi.org/10.5334/sta.bo.

Kraft, K., and W. D. Smart. 2016. "Seeing Is Comforting: Effects of Teleoperator Visibility in Robot-Mediated Health Care." In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 11–18. https://doi.org/10.1109/HRI.2016.7451728.

Mazur, Michael, and Adam Wisniewski. 2016. "Clarity from above. PwC Global Report on the Commercial Applications of Drone Technology." Price Water Cooper, PwC. https://www.pwc.pl/en/publikacje/2016/clarity-from-above.html.

Noddings, Nel. 2002. *Starting at Home: Caring and Social Policy*. Berkeley: University of California Press.

Poel, Ibo van de. 2013. "Why New Technologies Should Be Conceived as Social Experiments." *Ethics, Policy & Environment* 16 (3): 352–355. https://doi.org/10.1080/21550085.2013.844575.

Raymond, N., Brittany Card, and Ziad Al Achkar. 2012. "The Case Against Humanitarian Drones." *OpenCanada*, 2012. https://www.opencanada.org/features/the-case-against-humanitarian-drones/.

Reich, Warren T. 1995. "History of the Notion of Care." In *Encyclopedia of Bioethics*, edited by Warren T. Reich, 219–331. New York; London: Macmillan Pub. Co.: Simon & Schuster Macmillan: Prentice Hall International.

Sandvik, Kristin Bergtora, and Maria Gabrielsen Jumbert. 2016. *The Good Drone*. London; New York: Routledge Taylor & Francis Group.

Slim, Hugo. 2015. *Humanitarian Ethics: A Guide to the Morality of Aid in War and Disaster*. New York: Oxford University Press.

Soesilo, D., Patrick Meier, Audrey Lessard-Fontaine, Jessica Du Plessis, Christina Stuhlberger, and Valeria Fabbroni. 2016. "Drones in Humanitarian Action: Drones in Humanitarian Action." Swiss Foundation for Mine Action. https://reliefweb.int/report/world/drones-humanitarian-action-guide-use-airborne-systems-humanitarian-crises.

Tronto, Joan. 1993. *Moral Boundaries: A Political Argument for an Ethic of Care*. New York: Routledge.

Tronto, Joan. 2010. "Creating Caring Institutions: Politics, Plurality, and Purpose." *Ethics and Social Welfare* 4 (2): 158–171.

Tronto, Joan. 2013. *Caring Democracy: Markets, Equality, and Justice*. New York: NYU Press.

UN-OCHA. 2014. "Humanitarianism in the Age of Cyberwarfare Age: Towards the Principled and Secure Use of Information in Humanitarian Emergencies." OCHA Policy Paper. New York.

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.

Van De Poel, Ibo. 2018. "Moral Experimentation with New Technology." In *New Perspectives on Technology in Society: Experimentation beyond the Laboratory*, edited by Ibo van de Poel, Donna C. Mehos, and Lotte Asveld, 59–79. London: Taylor and Francis.

Vanlaere, L., and C. Gastmans. 2011. "A Personalist Approach to Care Ethics." *Nursing Ethics* 18 (2): 161–173.

van Wynsberghe, Aimee. 2013a. "Designing Robots for Care: Care Centered Value-Sensitive Design." *Science and Engineering Ethics* 19 (2): 407–433. https://doi.org/10.1007/s11948-011-9343-6.

van Wynsberghe, Aimee. 2013b. "A Method for Integrating Ethics into the Design of Robots." *Industrial Robot: An International Journal* 40 (5): 433–440. https://doi.org/10.1108/IR-12-2012-451.

van Wynsberghe, Aimee. 2016a. *Healthcare Robots: Ethics, Design and Implementation*. London: Taylor and Francis.

van Wynsberghe, Aimee. 2016b. "Service Robots, Care Ethics, and Design." *Ethics and Information Technology* 18 (4): 311–321. https://doi.org/10.1007/s10676-016-9409-x.

van Wynsberghe, Aimee, and Chris Gastmans. 2008. "Telesurgery: An Ethical Appraisal." *Journal of Medical Ethics* 34 (10), e22.

van Wynsberghe, Aimee, and Tina Comes. 2019. "Drones in Humanitarian Contexts, Robot Ethics, and the Human–Robot Interaction." *Ethics and Information Technology*, October. https://doi.org/10.1007/s10676-019-09514-1.

van Wynsberghe, Aimee, Denise Soesilo, Kristen Thomasen, and Noel Sharkey. 2018. "Drones in the Service of Society." Netherlands: Foundation for Responsible Robotics. https://responsiblerobotics.org/2018/06/05/report-drones-in-the-service-of-society/.

Verbeek, Peter-Paul. 2006. "Materializing Morality: Design Ethics and Technological Mediation." *Science, Technology, & Human Values* 31 (3): 361–380. doi:10.1177/0162243905285847.

Verbeek, Peter-Paul. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press, 2011.

Verkerk, Marian. 2001. "The Care Perspective and Autonomy." *Medicine, Health Care, and Philosophy* 4 (3): 289–294.

Watson, Jean. 1999. *Nursing: Human Science and Human Care: A Theory of Nursing*. Sudbury Mass.: Jones & Bartlett Learning.

Watson, Jean. 2011. *Human Caring Science*. Sudbury, MA: Jones & Bartlett Publishers.

Wilson, Melanie. 2002. "Making Nursing Visible? Gender, Technology and the Care Plan as Script." *Information Technology & People* 15 (2): 139–158.

CHAPTER 32

EMERGING TECHNOLOGY AS PROMISE AND PERIL

DEBORAH G. JOHNSON

1. Introduction

Technology has not traditionally been a major focus of philosophy, especially social and moral philosophy. While philosophers such as Plato and Aristotle, Francis Bacon, and Karl Marx discussed technology as part of their inquiries (Franssen, 2018), Heidegger's essay "The Question Concerning Technology" (1977 [1954]) is among the first and may indeed be the first work to place technology squarely in the purview of philosophy. Only in the decades after publication of Heidegger's essay (though not always drawing upon it), ground-breaking works on technology by philosophers began to appear. The writings of Hans Jonas, Don Ihde, Andrew Feenberg, and Albert Borgmann probed the social, political, and moral significance and hazards of modern technologies. In recent decades the field of philosophy of technology has grown and flourished, with a number of philosophers explicitly turning their attention to the connection between technology and "the good life." (See, for example, Higgs, Light and Strong 2010; Spence 2011; Swierstra and Waelbers 2012; and, Brey et al. 2012.)

The idea of "the good life" has a traditional philosophical meaning going back to Aristotle, for whom the good life and the virtuous life are one. Aristotle's notion of the good life is embedded in his account of human flourishing (*eudaimonia*) and the virtues. Although several contemporary philosophers of technology take an Aristotelian approach in thinking through the significance of technology (for example, Spence 2011, and Vallor 2016), most of the recent work uses the idea of the good life more broadly to refer to aspects of life that are considered valuable such as health, safety, convenience, and autonomy. In the Brey 2012 volume, for example, authors discuss well-being, quality of life, and happiness under the rubric of the good life.

An important aspect of this recent work is that it recognizes and builds on a relatively new understanding of technology that tightly connects technology to values and forms of living. The new view (which will be referred to here as the sociotechnical systems perspective) includes the claim that technology is more than just artifacts and the claim that technology shapes and is shaped by society. Although this understanding of technology may seem obvious now, it developed in the late 20th century as the field of science and technology studies (STS) developed and grew. The sociotechnical systems perspective rejects the idea that technology is just material objects, and the idea that technology is created exclusively by scientists and engineers working in isolation. The antiquated view as exemplified by authors such as Emmanuel Mesthene (1969) held that after having been created in isolation, technology is then delivered to society at which time decisions are made (by individuals and groups) about whether to accept and adopt or reject what has been delivered. The new, sociotechnical systems perspective acknowledges that new technologies develop in social, political, cultural and economic contexts. New technologies build on past technologies and scientific discoveries; nevertheless, a wide range of social factors influence what gets developed and when. Moreover, the sociotechnical systems perspective recognizes that technologies (especially complex, modern technologies such as railroads and automobiles) are ensembles of artifacts together with social practices, social relationships, institutional arrangements, and forms of knowledge. Technologies have material components, but they function only through a combination of machine and human behavior and through the meaning humans assign to them (Johnson and Wetmore 2009).

This means that technology does not just pervade human lives; it constitutes the world in which human beings live their lives and it is constituted by human practices, arrangements, and meaning. Technology constrains and enables what people do, organizes social relationships and institutions, and shapes individual and social values. That the connection between technology and society is not one-way is important to remember. Technology does not *just* shape the conditions of human lives and moral action, the technologies that come to be are themselves shaped by human practices, modes of living, and values. The way humans live at any given time influences the technologies that are developed. Technologies develop in particular social contexts with investors, government policies, cultural attitudes, historical events and more influencing the design and meaning of particular technologies.

A corollary to the sociotechnical systems perspective is the claim that technological change "in the making" today will have effects on human lives and values in the future. As technology and society evolve together through time, technological change today has implications not just for the immediate future but also for the longer-term future. It is this corollary that has led to an interest in anticipatory technology governance and ethics (Barben, Fisher, Selin and Guston 2008; Guston 2010, 2014; Brey 2012; Johnson 2011). Since technology and society are inextricably intertwined, we have the possibility of shaping society in the future by intentionally and conscientiously directing technological development today. By influencing the development of new technologies, we can influence the character of lives in the future. With the good life in mind, this means that we can through technology create conditions for good, or at least better, lives in the future.

This casts the relationship between technology and the future in a positive light. However, the relationship can also be framed in a threatening way, that is, if we don't carefully steer technological developments in the making today, we may degrade the possibilities for good lives in the future, that is, future generations may face diminished possibilities for good lives. This is the claim made by activists who are concerned about climate change and sustainability. Once the connection between technology and the character of human lives is made, however, the threat applies more broadly to most if not all technological decision making. This idea is evident in some recent warnings about technology. We are told, for example, that if we don't direct the development of artificial intelligence (AI) carefully, future generations may become utterly enslaved to machines and/or may not be able to think for themselves (Bostrom 2014).

The corollary points to the importance of identifying strategies for thinking about new and emerging technologies and their social and ethical implications while such technologies are still in the early stages of development. That is, anticipating the effects of new technologies while they are still in nascent form provides opportunities to direct their trajectories of development away from bad consequences and toward good effects.

One of the most commonplace modes of thinking about emerging technologies is in terms of their promises and perils. Futuristic imaginaries, predictions, and promotions of nascent technologies promise that this or that new technology will produce or improve upon some dimension of the good life, be it health, safety, or happiness. The promises are often, then, countered by claims (from those who are more cautious or skeptical or realistic) about the perils—diminished freedom, weakening of community ties, loss of privacy, new vulnerabilities to nature. A simple Google Scholar search reveals the commonplace use of promises and perils in recent scholarly literature on emerging technologies. Examples include: Heinrichs' "The Promises and Perils of Non-Invasive Brain Stimulation" (2012); Wolpe, Foster, and Lanleben's "Emerging Neurotechnologies for Lie-Detection: Promises and Perils" (2010); Nicholson's "The Promises and Perils of Geoengineering" (2013); and Eppler's "Information Quality and Information Overload: The Promises and Perils of the Information Age" (2015).

In certain respects, the promises-and-perils framework is an obvious and somewhat intuitive way of thinking about future technologies. Since future technologies do not yet exist, what else can we focus on except their potential? And that is precisely what promises and perils are, they point to potential benefits and potential harms. Perhaps because of its intuitive appeal, thinking in terms of promises and perils is rarely challenged; its value and limitations are rarely examined.

The aim of this chapter is to do just that—to consider the value and limitations of the promises-and-perils framework. Is it a worthy way of thinking about emerging technologies and the good life? What are its advantages and pitfalls? The analysis provided does not argue for abandoning the framework; it points to valuable as well as problematic dimensions of the framework. Most important, the analysis argues that a particular kind of critical perspective is essential when evaluating the plausibility of claims about emerging technologies.

2. Emerging Technologies and Promises and Perils

Emerging technologies are technologies that are acknowledged to be in the early stages of development. They don't yet exist in the sense that whatever their current state, those involved in the technology's development and use believe that what exists now is a nascent version of what it will be in its mature and stabilized form. Examples of such technologies include autonomous cars, humanoid robots, brain engineering, geoengineering, and artificial intelligence. Although models or versions of each of these are now available, current forms are understood to be embryonic forms of what will be available in the future.

On the one hand, because emerging technologies don't yet exist, the promises-and-perils framework seems appropriate. Promises and perils and emerging technologies are all about the future. Promises are declarations that something will happen or be done *in the future*; perils are harms or losses that may happen *in the future*; and, emerging technologies are touted not for what they are now but for what they will be *in the future*. The promises counterbalance the fact that hurdles have to be overcome before the vision of the technology can be fully realized; they justify investment of capital, effort, attention, and even hope. The perils argue for not investing or for caution in the pursuit or for a change in the apparent trajectory of development.

The hurdles that emerging technologies typically face are not just scientific and technological but also financial and social. For example, entrepreneurs trying to develop a new technology may have to convince investors, create a market, train workers and/or users, and garner public acceptance. Since no one knows for sure how and whether the hurdles to successful development and acceptance will be overcome, claims about the promises and perils of emerging technologies are always speculative. Uncertainty is inherent. Promises and perils are stories infused with assumptions and theories about how the world works and how technological change occurs. The stories come from a variety of actors—entrepreneurs, investors, and researchers, as well as potential users, government agencies, and journalists—each with different kinds of interests. Entrepreneurs and researchers tell stories to justify investment; users who stand to benefit or lose tell their stories about their needs; journalists are interested in garnering public interest and attention, and so on.

Because of the uncertainty inherent in promises and perils and the interests of those who promulgate them, a degree of skepticism is always appropriate. Any touted benefit or risk is speculation. Importantly, however, not all promises and perils are equal in their uncertainty. Some are more plausible than others. I will return to this point in the final section.

3. Promises and Perils and Cost-Benefit Analysis

Cost-benefit analysis (CBA) is an analytical tool for evaluating future endeavors. CBA seeks to identify and quantify the costs and benefits of undertaking any project that will have future consequences. The endeavor may be a policy change, a construction project, or investment in/adoption of a new technology. Arguably, CBA is a more fine-grained method compared to promises and perils. That is, CBA is more precise in requiring that consequences be specified in a quantifiable measure, typically dollar amounts, and the calculations are usually accompanied by a justification of how the calculations were made. CBA and promises-and-perils thinking are both strategies that have been directed at anticipating the effects of future technologies, and both have an underlying utilitarian structure.

CBA and promises-and-perils analyses can both be thought of as stories. They tell stories about what will happen in the future, and the stories either spell out or presume a sequence of events that will lead to future outcomes. Promises-and-perils stories tend to presume the sequence and project farther into the future while CBA stories tend to be more explicit about assumptions because the assignment of costs and benefits requires justification. In promises about emerging technologies the story sequence typically presumes that technological hurdles will be overcome, particular policy frameworks will be adopted, regulators' approvals will be obtained, and the public will accept a particular design and reconfigured social arrangements. Peril stories typically presume or explicitly identify sequences in which the development process misses some negative consequence entirely or actors with heinous motivations come to control the development process or some unexpected event occurs that no one anticipated. Examples of promise or peril stories that implicitly presume a sequence of events that will move an emerging technology from its present state to a future state are not hard to find. For example, the promise of autonomous cars safer than human-driven cars presumes many technological hurdles being overcome, particular insurance and standards setting regimes being adopted, and the public becoming comfortable putting their lives in the hands of an autonomous machine. The sequence is not spelled out in detail; promise stories simply presume that these events will occur. The same can be said for peril stories. Consider the peril of robots becoming so sophisticated and so much more intelligent than humans that they organize and rebel. Stories of this kind assume that the enormous technological challenges of achieving artificial intelligence comparable to human intelligence will be overcome and that those involved in this development will fail to anticipate and design to prevent robot rebellion.

Although the stories that CBA and promises-and-perils thinking tell about emerging technologies can be quite different, there is an abundant literature focusing on CBA and its weakness and little critical examination of promises-and-perils thinking. The criticisms of CBA provide a useful basis for teasing out some of the advantages and disadvantages of promises-and-perils thinking. That is, we can ask how the promises-and-perils framework fares in relation to the standard criticisms of CBA.

3.1 Overestimating Benefits/Underestimating Costs

CBA is often criticized for its tendency to overestimate benefits and underestimate costs (Boardman 2017; Flyvbjerg, Holm and Buhl 2003). Especially when those doing the analysis have an interest in going forward with an endeavor, good consequences can be given higher dollar values than they will ultimately have, and bad consequences can be assigned lower values. For example, contractors who use CBA to convince clients of the value of going forward with construction projects have an obvious interest in embellishing the benefits and downplaying the costs. They want the client to say "yes" to going forward with the project.

This tendency of CBA is also a vulnerability of promise-and-peril thinking. Indeed, the potential for embellishing and/or discounting promises and perils seems an intractable problem. Uncertainty about future consequences can easily be exploited by those who have an interest in a nascent technology's development or non-development. For example, entrepreneurs and researchers may overstate the promises and downplay the perils (the risks) of geoengineering as a viable approach to global warming because they want to convince investors and sponsors to fund their research; medical professionals or military representatives who see their interests served by brain engineering may exaggerate the promise and not acknowledge the peril of undermining notions of what it means to be a person. It is the same for perils, as in the case of unions criticizing anticipated AI technologies as likely to replace workers in the future. Because promises and perils are by nature less formal and make implicit presumptions about the sequence of events that will get to the promise or peril, the potential for embellishment is even greater than with CBA. Promises-and-perils thinking is not typically burdened with the requirement of rigorous justification of claims.

3.2 Nonquantifiable Costs and Benefits

CBA is also criticized for its inability to deal with non-quantifiable costs and benefits. One of the advantages of CBA is that costs and benefits are formulated in a single measure. This allows for a precise balancing. As CBA advocates would have it, trade-offs are difficult to make when one is dealing with "apples and oranges" but when costs and benefits are all translated into a single measure, the balancing and trade-offs are much clearer. As already mentioned, in CBA, the single measure is typically monetary, the dollar value of each benefit and cost. This, then, leads to the criticism that many of the positives and negatives of a future technology can't be adequately translated into monetary amounts or any other quantitative measure. The most obvious example is the value of a human life. In the infamous Pinto case, Ford Motor Company was severely criticized for putting a dollar value on human life when it calculated that it would be cheaper for the company to pay off an estimated number of legal suits for loss of a life than to pay for a part that would have made the Pinto safer, i.e., less likely to explode in rear-end collisions (De George 1981; May 1982). CBA seems ill-equipped to cope with many non-quantifiable consequences, not just the value of human life. How can a dollar value be assigned to personal privacy, free speech, democratic institutions, or national cohesion?

This criticism of CBA does not apply to promises and perils thinking because promises and perils can be, and often are, expressed in non-quantitative terms. Of course, it depends on the technology and the actors who are identifying the promises and/or perils, but in general the promises and perils of new technologies are commonly discussed in non-quantitative terms. In the case of autonomous cars, both strategies are found in the literature. CBA has been used to calculate the monetary costs of a certain number of autonomous automobile accidents including the costs of deaths and injuries, as well as the costs of traffic congestion and parking spaces. These costs are then juxtaposed against the cost of an estimated reduction in the number of automobile accidents, amount of traffic congestion, and need for parking spaces were a system of autonomous cars in operation (See KPMG 2012 and Fagnant and Kockelman 2015). On the other hand, there has been broader discussion of the promises and perils of autonomous cars in non-quantitative terms including the promise of easier access to transportation for the disabled, the perils of a reduction in individual freedom (or at least the feeling of freedom) from not being able to drive, and a reduction in privacy since the operations of autonomous cars will be data intensive and may even require cameras inside the cars (Kent 2014; Boeglin 2015). These are non-quantifiable benefits and costs.

Humanoid social robots are an example of an emerging technology for which promises and perils seem better suited than CBA. To be sure, CBA can and has been used for industrial robots, but the adoption of robots that increasingly look like humans for use as companions to children and the elderly or even for use as sexual partners seems better captured in non-quantifiable promises and perils than in monetary costs and benefits (Sharkey and Sharkey 2011; Levy 2009).

3.3 Distribution of Costs and Benefits

CBA is often criticized for its inability to account for the uneven distribution of costs and benefits. This is a problem it inherits from utilitarianism. The strict focus on costs and benefits disregards the fact that costs and benefits often affect individuals and groups unevenly. New technologies affect different individuals and groups differently; some benefit while others are put at increased risk; some pay the cost while others are made better off. In fact, CBA often works to make the worse-off even more so. As is well-known, when CBA is used to calculate the best place to site new manufacturing or waste disposal plants—plants that produce toxins and pollute the air or water near the plant—the plants end up in the poorest neighborhoods. CBA tends to recommend low income neighborhoods because of the low real estate values and because in these places, companies have leverage to negotiate reduced taxes because the community so desperately needs jobs (Bullard 2018). The uneven distribution of the benefits and burdens of emerging technologies is also a global problem in that new technologies often benefit those living in wealthy industrialized societies while burdening those in less wealthy regions of the world. A good example of this is e-waste. When cellphones and computers are used and then discarded in the United States, they are shipped to less wealthy nations such as China, India, and Africa where recycling methods expose workers (including children) and those living nearby to dangerous chemical toxins (Whitehouse 2012). CBA might show that the benefits of electronic products far outweigh the costs, but the calculation does not take into account that the costs and benefits are borne by very different groups and make the already worse off more so.

To be sure, there are promise-and-peril discourses on emerging technologies that also ignore distributional effects or that do not consider effects on certain groups such as women and minorities when they imagine future technologies. Even when it comes to perils, much of the discourse drawing attention to massive job losses resulting from AI ignores how those job losses will disproportionately impact some groups more than others. AI controlled robots are more likely to replace lower skilled jobs and since women hold a disproportionate number of lower skilled jobs, women will be affected more severely (Brinded 2017).

Both CBA and promises-and-perils thinking are, then, vulnerable to this neglect of distributional issues, but CBA is structurally unable to handle distributive issues effectively, while promises-and-perils thinking can easily accommodate attention to distributive issues because it can use qualitative terms, such as justice, fairness, and equality. In fact, in the current discourse on geoengineering, the differential regional impacts have received a fair amount of attention as one of the perils that has not been recognized by CBA (Heyen, Wiertz and Irvine 2015; Tuana et al. 2012).

3.4 Unforeseen and Unforeseeable Consequences

Finally, CBA is criticized for not being able to take unforeseen and unforeseeable consequences into account in balancing costs and benefits. This is problematic not just because it is impossible to assign a dollar value to what isn't seen; the problem is that such consequences are "unforeseen and unforeseeable." However, that there can or will be such consequences is specifiable and often addressed in the promises and perils framework. The unforeseen is often put forward as one of the perils, that is, one of the reasons to be concerned about going forward with a new technology. Hansson (2005) defends the significance of this type of peril broadly: "if someone proposed to eject a chemical substance into the stratosphere for some good purpose or other, it would not be irrational to oppose this proposal solely on the ground that it may have unforeseeable consequences . . . " (p. 75). Heyward (2014) is more measured in noting only that unforeseeable consequences must be addressed in the context of geoengineering technologies:

"If societies are ever to make an assessment of the proper place of CDR [carbon dioxide removal]and SRM [solar radiation management] technologies in any portfolio of responses to anthropogenic climate change, the issue of unforeseeable harms will have to be dealt with" (p. 406).

So, when it comes to unforeseen and unforeseeable consequences, the promises and perils framework seems to have an advantage over CBA. Of course, the framework cannot identify the unidentifiable, but the framework allows the unforeseen to be part of the consideration as to whether to go forward with a new technology.

In short, then, when promises-and-perils thinking is evaluated in relation to the standard criticisms of CBA, some of its strengths and weaknesses are revealed. Promises-and-perils thinking seems no better or worse than CBA in its potential to overestimate and underestimate the future consequences of an emerging technology. On the other hand, promises and perils thinking is not burdened with the requirement of putting costs and benefits into a quantifiable form and this seems an advantage over CBA. That is, promises and perils thinking is better equipped to draw attention to non-quantifiable benefits and costs. Promises and perils thinking also seems better equipped to draw attention to the uneven distribution of benefits and burdens of emerging technologies, though there is nothing in this way of thinking that guarantees distributive issues will be addressed. Finally, as with the distributive issue, promises and perils thinking is better equipped to deal with the unforeseeable, not because it allows the unforeseeable to be seen, but because it can make a place for this in the discourse while CBA has no way of dealing with it.

4. The Uncertainty and Plausibility of Promises and Perils

The overarching problem with promises and perils, implicit in what has just been discussed, is its speculative nature. Whether or not a promise or peril will ever be realized is uncertain. An emerging technology might or might not develop as projected and it might or might not lead to the promised or perilous outcome. Claims about the future are unverifiable in the present. Even with a simple promise from one person to another, there is uncertainty as to whether the promiser will do what they promise. Perils are situations involving risk of injury, harm, loss or destruction, but whether these risks will become realities is not certain. Uncertainty about the nature of future technologies means uncertainty in the identification of consequences, the assignment of values to possible consequences, accounting for how consequences will be distributed, and addressing unforeseen consequences.

However, even though uncertainty is inherent in promises-and-perils thinking, it would be a mistake to accept uncertainty *writ large*, as if all promises and perils were equal. Promises and perils are more and less plausible and, therefore, more and less valuable for decision-making about emerging technologies. Some are ludicrous and fantastical; some reflect the interests of those who tout them; some are based on lessons learned from other technologies; and so on. For this reason, identifying a basis for evaluating promises and perils is critically important. Recent work has taken up the broad challenge of understanding plausibility in futuristic claims though the matter is far from well-understood especially when it comes to claims about emerging technologies (Selin 2011; Selin and Pereira 2013).

Two caveats are worth mentioning here. First, uncertainty is not unique to promises-and-perils thinking; it is inherent in other anticipatory endeavors such as scenario development and foresight tools. Indeed, it is inherent to any futuristic planning strategy. Second, narratives about emerging technologies are motivated by varying interests and serve a variety of purposes including entertainment and social commentary, and for some of these purposes, plausibility is not important. For example, "Sultana's Dream" by Rokeya Sakhawat Hossain (2005 [1905]) depicts fantastical technologies but the thrust of the story is a powerful feminist critique.

Nevertheless, there are a host of promises and perils in the discourse around emerging technologies for which plausibility is important. Should I accustom myself to the idea that in the future autonomous cars may well be safer than driving myself? Should government funding be allocated to geoengineering as a plausible approach to climate change? Should a young person avoid certain career trajectories because AI is likely to eliminate those careers in the future? Should I make financial investments in 5G technology? Answering these kinds of questions requires some sort of evaluation of the plausibility of touted promises and perils.

Although a comprehensive account of plausibility will not be provided here, one important piece of such an account is provided by acknowledging the sociotechnical nature of technology. This perspective enables identification of multiple sources of uncertainty all of which should be acknowledged. Promises and perils that focus exclusively on the material form (the artifact) of an emerging technology are less plausible than those that include the social arrangements and practices and social meanings that will constitute a future technology.

Although this makes anticipatory endeavors complex—because of the multiplicity of diverse factors that can influence technological development—we can (for heuristic purposes at least) distinguish three sources of uncertainty. The first has to do with the features of the artifactual components of the technology and what form the artifact might have in the technology's mature, stabilized state. The second source of uncertainty has to do with the characteristics of the system in which the artifactual components will operate, and the third has to do with the social meaning and consequences of adopting a technology with particular artifactual and system features.

Of course, these three uncertainties are not manifested separately or sequentially. The artifactual features, system features, and social context and meaning of an emerging technology interact with each other throughout the development process. Each dimension is worked out through negotiations with the others. The artifact may change as system issues arise; the system may change as artifactual challenges are overcome in particular ways; the design of the artifact and/or the system may change as social attitudes become evident and suggest better public acceptance if this or that change is made in the artifactual or system design. Developers are simultaneously directing engineers to make prototypes, searching for capital, talking to regulators, interviewing potential consumers and these all affect how a technology moves from its nascent form to a successful, stabilized form.

Some might argue that when it comes to emerging technologies, there isn't all that much uncertainty about the artifact because the very idea of a particular, new technology indicates features from which one can project into the future. In the case of autonomous cars, the very idea of cars without human drivers gives some basis for identifying promises and perils. The same argument could be made about nanotechnology for which the basic idea is very small-scale manipulation of materials or about brain engineering for which the simple idea is altering brain cells, and so on. The basic idea of the artifact provides grounds for speculation about its possible benefits (promises) and possible dangers (perils). This is true. Nevertheless, the lack of specificity as to how the simple idea will be achieved greatly limits, and can even misdirect attention as to, the validity of touted promises or perils.

The three sources of uncertainty can be illustrated by considering the discourse around autonomous cars. The overarching promise trumpeted in the discourse on autonomous cars is safety. The promise is not just that the cars will be safe but that they will be safer than human-driven cars. Autonomous cars are promoted as leading to fewer automobile accidents and, hence, fewer fatalities and injuries. Elon Musk has touted the promise of safety to such an extent that he has claimed that the Tesla would soon become so safe that having a human intervene will actually decrease safety (Tung 2019). Yet determining whether autonomous cars will ever be safe enough (let alone safer than driver cars) does not just depend on whether and how engineers will solve a whole host of artifactual challenges. It depends on overcoming the challenges of building a system—an infrastructure—in which autonomous cars can operate safely, and also on whether the public can be convinced to trust the cars and give up driving. Influencing all of these challenges is answering the key question as to how safety will be determined, measured and certified in autonomous cars. This is not a purely technical matter. The idea of a car that operates without a driver is fanciful unless there is some promise of safety, and the promise of safety depends on much more than overcoming artifactual design.

To be sure, autonomous car developers claim to have figured out the main types of artifactual components that will be necessary to make the cars work safely. According to one account, it is a matter of four components: sensing (radar, lidar, etc.), detection (detecting objects, etc.), perception (environment modeling, localization, etc.), and decision making (obstacle avoidance, prediction, etc.; Sovani 2018). Many of the proclaimed promises of autonomous cars are based on projections about how these technical challenges will be met.

Nevertheless, autonomous cars will operate in a larger technological system as well as economic, political and legal systems. Whether the promise of safety (or of such perils as massive unemployment from autonomous cars) comes to be is largely dependent on the kind of systems in which the cars will operate. For example, will there be a system of purely autonomous cars or a system in which autonomous and human-driven cars operate alongside each other? For safety, the pure system will be easier to achieve; however, the transition to only autonomous cars is a daunting social and economic challenge. Will all car owners in one-fell-swoop turn in their cars in exchange for autonomous ones? If not, what will happen during the transition? Who will be affected the most and least by a transition period, especially in economies where driving is a primary source of well-paying jobs for those without higher education? Will autonomous cars be owned by individuals or will the cars be owned by private companies or public organizations such as cities and called up by individuals when they need them? How will insurance be handled? Of course, the insurance schema will depend on the ownership schema and the cost. All of this is to say that promises and perils that do not take the system into account, e.g., those that simply project an artifact with particular features, are much less reliable than those that do.

Typically, those who are focused on the artifact make assumptions about the system in which the cars will operate, but such assumptions are typically made in an ad hoc manner. Yet, the system issues are as essential as the artifactual to whether promises and perils are plausible.

The third source of uncertainty has to do with how people will think about and be affected by the technology. The sociotechnical system perspective keeps in the forefront that autonomous cars will function not just by means of artifactual components and systems; they will also be shaped by the meaning that human beings assign to them and the ways in which people relate to them. People will interact with the cars as passengers, pedestrians, possibly owners, but also people will monitor the cars and the system, and will be involved in the manufacture, marketing, distribution, and maintenance of the cars and the system. People will also make personal decisions about safety, decisions about when, if ever, the cars are safe enough to be put on public roads. For example, some have argued that autonomous cars will have to be four to five times safer than human-driven cars before the public will accept them (Liu, Yang and Xu 2019). The point is that whether or not the touted promises and perils of autonomous cars are realized depends on how people respond. Although research continues to be done on this, there is still a good deal of uncertainty about how humans will feel about being transported in a driverless vehicle.

Acknowledging these three different kinds of uncertainty provides a basis for evaluating touted promises and perils. Promises and perils that are focused exclusively on the artifactual features of an emerging technology, and make ad hoc assumptions about the technological and social systems in which the artifact will operate, are not reliable. Promises and perils that acknowledge that in order to successfully mature, an emerging technology will have to fit into a broader sociotechnical system and acquire social meaning and acceptance are much more reliable. The more a promise or peril story addresses all three kinds of uncertainty, the more plausible it is.

5. Conclusion

Where does this leave us with respect to the promises-and-perils framework as an anticipatory approach to emerging technologies? Because technologies co-constitute the world we live in and new technologies may reconstitute and reconfigure that world for good or ill, anticipating the consequences of emerging technologies is imperative. The promises-and-perils framework has an intuitive appeal and has certain advantages insofar as it can accommodate nonquantifiable considerations, the distribution of consequences, and the possibility of unforeseen and unforeseeable consequences. The primary problem with promises and perils is their uncertainty and the challenge of evaluating the plausibility of narratives in which they are embedded. This is a problem that the promises-and-perils framework shares with other anticipatory and futuristic endeavors.

Because claims about the promises and perils of emerging technologies are inherently uncertain, evaluation of the plausibility of such claims is important. Although this chapter has not provided a comprehensive approach to this evaluation, it has argued that claims about the promises and perils of emerging technologies can be distinguished by the extent to which they acknowledge three different sources of uncertainty in technological development. Promises and perils intended to influence decision-making about emerging technologies should not just focus on artifactual design. In order to be plausible, they should address the broader technological system in which the artifact will operate as well as the social context and social meaning of the technology.

References

Barben, Daniel, Erik Fisher, Cynthia Selin, and David H. Guston. 2008. "Anticipatory Governance of Nanotechnology: Foresight, Engagement, and Integration." In *The Handbook of Science and Technology Studies,* edited by Edward Hackett, Olga Amsterdamska, Michael Lynch, and Judy Wajcman, 979–1000. Cambridge: MIT Press.

Boardman, Anthony E., David H. Greenberg, Aidan R. Vining, and David L. Weimer. 2017. *Cost-Benefit Analysis: Concepts and Practice*. Cambridge University Press.

Boeglin, Jack. 2015. "The costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in autonomous Vehicle Regulation." *Yale Journal of Law & Technology* 17, no. 1: 171–303.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Brey, Philip, Adam Briggle, and Edward Spence, eds. 2012. *The good life in a technological age*. Routledge.

Brey, Philip AE. 2012. "Anticipatory ethics for emerging technologies." *NanoEthics* 6, no. 1: 1–13.

Brinded, Lianna. 2017. "Robots Are Going to Turbo Charge One of Society's Biggest Problems." *Quartz*, December 28, 2017. https://qz.com/1167017/robots-automation-and-ai-in-the-workplace-will-widen-pay-gap-for-women-and-minorities/

Bullard, Robert D. 2018. *Dumping in Dixie: Race, Class, and Environmental Quality*. 3rd edition. Routledge.

De George, Richard T. 1981. "Ethical Responsibilities of Engineers in Large Organizations: The Pinto Case." *Business & Professional Ethics Journal* 1, no. 1: 1–14.

Eppler, Martin J. 2015. "Information Quality and Information Overload: The Promises and Perils of the Information Age." In *Communication and Technology,* edited by Lorenzo Cantoni and James A. Danowski, 215–232. Walter de Gruyter GmbH.

Fagnant, Daniel J., and Kara Kockelman. 2015. "Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations." *Transportation Research Part A: Policy and Practice* 77: 167–181.

Flyvbjerg, Bent, Mette K. Skamris Holm, and Søren L. Buhl. 2003. "How Common and How Large Are Cost Overruns in Transport Infrastructure Projects?" *Transport Reviews* 23, no. 1: 71–88.

Franssen, Maarten, Gert-Jan Lokhorst, and Ibo van de Poel. 2018. "Philosophy of Technology." In *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/fall2018/entries/technology/

Guston, David H. 2010. "The Anticipatory Governance of Emerging Technologies." *Journal of the Korean Vacuum Society* 19, no. 6: 432–441.

Guston, David H. 2014. "Understanding 'Anticipatory Governance.' " *Social Studies of Science* 44, no. 2: 218–242.

Hansson, Sven Ove. 2005. "The Epistemology of Technological Risk." *Techné* 9, no. 2: 68–80.

Heidegger, Martin. 1977 [1954]. The Question of Technology. *Vorträge und Aufsätze*. Garland Publishing.

Heinrichs, Jan-Hendrik. 2012. "The Promises and Perils of Non-invasive Brain Stimulation." *International Journal of Law and Psychiatry* 35, no. 2: 121–129.

Heyen, Daniel, Thilo Wiertz, and Peter James Irvine. 2015. "Regional Disparities in SRM Impacts: The Challenge of Diverging Preferences." *Climatic Change* 133, no. 4: 557–563.

Heyward, Clare. 2014. "Benefiting from Climate Geoengineering and Corresponding Remedial Duties: The Case of Unforeseeable Harms." *Journal of Applied Philosophy* 31, no. 4: 405–419.

Higgs, Eric, Andrew Light, and David Strong, eds. 2010. *Technology and the Good Life?* University of Chicago Press.

Hossain, Rokeya Sakhawat. 2005 [1905]. *Sultana's Dream; And Padmarag: Two Feminist Utopias*. Penguin Books India.

Johnson, D. G., and J. M. Wetmore, eds. 2009. *Technology & Society: Engineering Our Sociotechnical Future*. Cambridge: MIT Press.

Johnson, Deborah G. 2011. "Software Agents, Anticipatory Ethics, and Accountability." In *The Growing Gap between Emerging Technologies and Legal-Ethical Oversight*, edited by Gary E. Marchant, Braden R. Allenby, and Joseph R. Herkert, 61–76. Dordrecht: Springer.

Kent, Jennifer L. 2014. "Driving to Save Time or Saving Time to Drive? The Enduring Appeal of the Private Car." *Transportation Research Part A: Policy and Practice* 65: 103–115.

KPMG. 2012. "Self-Driving Cars: The Next Revolution." https://institutes.kpmg.us/manufacturing-institute/articles/2017/self-driving-cars-the-next-revolution.html

Levy, David. 2009. *Love and Sex with robots: The evolution of Human-Robot Relationships*. New York: Harper Collins e-books.

Liu, Peng, Run Yang, and Zhigang Xu. 2019. "How Safe Is Safe Enough for Self-Driving Vehicles?." *Risk Analysis* 39, no. 2: 315–325.

May, William W. 1982. "$ s for Lives: Ethical Considerations in the Use of Cost/Benefit Analysis by For-Profit Firms." *Risk Analysis* 2, no. 1: 35–46.

Mesthene, Emmanuel. 1969. "Some General Implications of the Research of the Harvard University Program on Technology and Society." *Technology and Culture* 10, no. 4: 489–513.

Nicholson, Simon. 2013. "The Promises and Perils of Geoengineering." In *State of the World 2013*, 317–331. Washington, D.C.: Island Press.

Selin, Cynthia. 2011. "Negotiating Plausibility: Intervening in the Future of Nanotechnology." *Science and Engineering Ethics* 17, no. 4: 723–737.

Selin, Cynthia, and Ângela Guimarães Pereira. 2013. "Pursuing Plausibility." *International Journal of Foresight and Innovation Policy* 9, no. 2-3-4: 93–109.

Sharkey, Amanda, and Noel Sharkey. 2011. "Children, the Elderly, and Interactive Robots." *IEEE Robotics & Automation Magazine* 18, no. 1: 32–38.

Sovani, Sandeep. 2018. "Top 3 Challenges to Produce Level 5 Autonomous Vehicles." *Ansys Blog*. https://www.ansys.com/blog/challenges-level-5-autonomous-vehicles

Spence, Edward H. 2011. "Is Technology Good for Us? A Eudaimonic Meta-Model for Evaluating the Contributive Capability of Technologies for a Good Life." *NanoEthics* 5, no. 3: 335–343.

Swierstra, Tsjalling, and Katinka Waelbers. 2012. "Designing a Good Life: A Matrix for the Technological Mediation of Morality." *Science and Engineering Ethics* 18, no. 1: 157–172.

Tuana, Nancy, Ryan L. Sriver, Toby Svoboda, Roman Olson, Peter J. Irvine, Jacob Haqq-Misra, and Klaus Keller. 2012. "Towards Integrated Ethical and Scientific Analysis of Geoengineering: A Research Agenda." *Ethics, Policy & Environment* 15, no. 2: 136–157.

Tung, Liam. 2019. "Elon Musk on Tesla's Autopilot: In a Year, 'A Human Intervening Will Decrease Safety.'" ZDNet. April 2019. https://www.zdnet.com/article/elon-musk-on-teslas-autopilot-in-a-year-a-human-intervening-will-decrease-safety/

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.

Whitehouse, Timothy. 2012. "E-Waste Exports: Why the National Strategy for Electronics Stewardship Does Not Go Far Enough." *George Washington Journal of Energy & Environmental Law* 3, no. 1: 110–116.

Wolpe, Paul Root, Kenneth R. Foster, and Daniel D. Langleben. 2010. "Emerging Neurotechnologies for Lie-Detection: Promises and Perils." *The American Journal of Bioethics* 10, no. 10: 40–48.

Index