

SPACES FOR THE FUTURE

A Companion to Philosophy of Technology

EDITED BY JOSEPH C. PITT
AND ASHLEY SHEW

Spaces for the Future

Focused on mapping out contemporary and future domains in philosophy of technology, this volume serves as an excellent, forward-looking resource in the field and in cognate areas of study. The 32 chapters, all of them appearing in print here for the first time, were written by both established scholars and fresh voices. They cover topics ranging from data discrimination and engineering design, to art and technology, space junk, and beyond. *Spaces for the Future: A Companion to Philosophy of Technology* is structured in six parts: (1) Ethical Space and Experience; (2) Political Space and Agency; (3) Virtual Space and Property; (4) Personal Space and Design; (5) Inner Space and Environment; and (6) Outer Space and Imagination. The organization maps out current and emerging spaces of activity in the field and anticipates the big issues that we soon will face.

Joseph C. Pitt, a Fellow of AAAS, has been teaching at Virginia Tech for forty-five years. His major area of scholarship is philosophy of technology with an emphasis on the impact of technological innovation on scientific change. He is the author of four books and the editor/co-editor of twelve others. He is past president of the Society for Philosophy and Technology and served as editor-in-chief of *Techné: Research in Philosophy and Technology*, of which he is currently on the editorial board.

Ashley Shew works in philosophy of technology on topics relating to emerging technologies, animals, disability, and environment. Her first book, *Animal Constructions and Technological Knowledge*, will be published next year. Her latest work concerns how disabled bodies are used in technological imagination and how narratives about technology often fail to match up to the lived experience of disability. She collects class materials and information about disability and technology on her website, <http://techanddisability.com>.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Spaces for the Future

A Companion to Philosophy
of Technology

Edited by Joseph C. Pitt
and Ashley Shew

First published 2018
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2018 Taylor & Francis

The right of Joseph C. Pitt and Ashley Shew to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book has been requested

ISBN: 978-0-415-84296-9 (hbk)
ISBN: 978-0-203-73565-7 (ebk)

Typeset in Times New Roman
by Apex CoVantage, LLC

Visit the companion website: <http://spacesforthefuture.com>

Contents

<i>List of Contributors</i>	viii
<i>Editor Introduction</i>	xii

PART I	
Ethical Space and Experience	I
1 Data, Technology, and Gender: Thinking About (and From) Trans Lives	3
ANNA LAUREN HOFFMANN	
2 Discrimination	14
D. E. WITTKOWER	
3 Video Games and Ethics	29
MONIQUE WONDERLY	
4 Social Networking: The Dialectics of Sharing	42
YONI VAN DEN EEDE AND KATLEEN GABRIELS	
5 Technoart	54
DON IHDE	
6 The Philosophy of Art and Technology	62
JESSIE MANN	
PART 2	
Political Space and Agency	71
7 Robots With Guns	73
SHANNON VALLOR	
8 Educational Technology	82
ANDREW WELLS GARNAR	

9 Religious Transcendence	92
CARL MITCHAM	
10 Agency and Citizenship in a Technological Society	98
ANDREW FEENBERG	
11 Animals in Philosophy of Technology	108
ASHLEY SHEW	
 PART 3	
Virtual Space and Property	117
 12 Obscurity and Privacy	119
EVAN SELINGER AND WOODROW HARTZOG	
 13 Copyright Between Economic and Cultural Models of Creativity	130
GORDON HULL	
 14 Cyberwarfare	141
BENJAMIN C. JANTZEN	
 15 The Cloud	154
CRAIG A. CONDELLA AND JULIE C. SWIERCZEK	
 16 Ethical Issues in Big Data	164
DEBORAH G. JOHNSON	
 17 Virtual Environments	174
JOHNNY HARTZ SØRAKER	
 PART 4	
Personal Space and Design	183
 18 The Organization of User Experience	185
ROBERT ROSENBERGER	
 19 Engineering Design	196
PIETER E. VERMAAS	
 20 Design	208
ANN JOHNSON	
 21 Dancing the Device: A Translational Approach to Technology	216
ALBRECHT FRITZSCHE	

22 On Harnessing of Birth in the Technical Age	224
DANA S. BELU	
23 Moore's Regula	238
CYRUS C. M. MODY	
 PART 5	
Inner Space and Environment	249
24 Is Renewable Energy Technology Revolutionary?	251
ROBERT DARROW	
25 Geoengineering and Climate Change	263
TINA SIKKA	
26 Fracking	281
ADAM BRIGGLE	
27 Climate Change and Philosophy of Technology	292
ROBERT-JAN GEERTS	
 PART 6	
Outer Space and Imagination	303
28 Space Tourism and Science Fiction	305
ROSALYN W. BERNE	
29 Well-Ordered Engineering: Participatory Technology Assessment at NASA	314
ZACHARY PIRTLE AND DAVID C. TOMBLIN	
30 When Loving Your Planet Isn't Enough	326
DIANA P. HOYT	
31 Transcendence in Space	340
JOSEPH C. PITT	
32 The Role of Technology in Natural Science	346
NICHOLAS RESCHER	
 <i>Index</i>	353

Contributors

Dana S. Belu is Associate Professor of Philosophy in the Philosophy Department at California State University, Dominguez Hills. She works at the intersection of phenomenology, philosophy of technology, and feminist philosophy. Her book *Heidegger, Reproductive Technology, and the Motherless Age* came out in 2017.

Rosalyn W. Berne explores the intersecting realms of emerging technologies, science, fiction, and myth. She is author of *Nanotalk: Conversations with Scientists and Engineers About Ethics, Meaning and Belief in the Development of Nanotechnology* (2005), *Creating Life From Life: Biotechnology and Science Fiction* (2014), and the science fiction novel *Waiting in the Silence* (2012).

Adam Briggie is Associate Professor in the Philosophy and Religion Department at the University of North Texas. He is the author of *A Field Philosopher's Guide to Fracking* (2015) and co-author with Robert Frodeman of *Socrates Tenured: The Institutions of 21st Century Philosophy* (2016).

Craig A. Condella is Associate Professor in the departments of Philosophy and of Cultural, Environmental, and Global Studies at Salve Regina University. His interests include the philosophy of technology, environmental ethics, and the history and philosophy of science.

Robert Darrow is a PhD Candidate in Political Science at University of Massachusetts, Amherst, and holds a master's degree in Philosophy from Virginia Tech. He is currently researching the large-scale adoption of renewable energy technology in Denmark. He has also written about mountaintop-removal coal mining in Appalachia.

Andrew Feenberg is Canada Research Chair in Philosophy of Technology at Simon Fraser University, and Directeur de Programme in the College International de Philosophie in Paris. He is the author of *Questioning Technology* (1999), *Between Reason and Experience* (2010), and *The Philosophy of Praxis; Marx, Lukács, and the Frankfurt School* (2014).

Albrecht Fritzsche works at the Institute of Information Systems of the University of Erlangen-Nuremberg. He holds doctoral degrees in philosophy and management and worked for many years in the manufacturing industry. His research is focused on innovation studies.

Katleen Gabriels is Assistant Professor at Eindhoven University of Technology in the Netherlands. Her research focuses on the ethical aspects of the Internet of Things and the virtualization of morality. She publishes within philosophy and ethics of technology and media studies.

Andrew Wells Garner is Senior Lecturer in Clemson University's Department of Philosophy and Religion. He teaches classes on science, technology, and postmodernism. His research deals with these topics as well, from the perspective of American pragmatism.

Robert-Jan Geerts is Assistant Professor in Ethics and Technology at University of Twente. He is interested in the interface between philosophy of technology and environmental philosophy, and his work explores avenues to broaden debates on energy transition.

Woodrow Hartzog is the W. Stancil Starnes Professor of Law at Samford University. His research focuses on privacy law, media law, internet law, intellectual property, torts, and contracts. Currently he is writing *Privacy's Blueprint: The Battle to Control the Design of New Technologies*.

Anna Lauren Hoffmann is Assistant Professor at the Information School at the University of Washington. Her research focuses on information, technology, culture, and ethics, with particular focus on how design and discourses of information technology can promote or hinder social justice.

Diana P. Hoyt is a senior policy analyst at National Aeronautics and Space Administration (NASA) Headquarters and currently manages NASA's Regional Economic Development and Innovation program. Her primary research interest is the philosophy of technology.

Gordon Hull is Associate Professor of Philosophy and Public Policy at the University of North Carolina, Charlotte, where he directs the Center for Professional and Applied Ethics. He works on topics in law, technology, and political philosophy such as intellectual property and privacy.

Don Ihde is Distinguished Professor of Philosophy, Emeritus, at Stony Brook University, and author of twenty-two books, most recently *Acoustic Technics* (2015) and *Husserl's Missing Technologies* (2016).

Benjamin C. Jantzen joined the Philosophy Department at Virginia Tech in 2011. He works broadly in the philosophy of science and the logic of discovery. Currently, he is developing computer algorithms for the automated discovery of scientifically relevant natural kinds.

Ann Johnson passed away before the completion of this volume and is dearly missed by her colleagues in history and philosophy of engineering. She did qualitative research on engineers and engineering, focusing particularly on the way engineers produce and circulate new knowledge and the role of specialized communities in the dissemination of knowledge. Her book, *Hitting the Brakes* (2009), is a study of knowledge production and community formation in the design of antilock braking systems for passenger automobiles.

Deborah G. Johnson is the Anne Shirley Carter Olsson Professor of Applied Ethics, Emeritus, in the Science, Technology, and Society Program at the University of Virginia. Best known for her work on computer ethics and engineering ethics, Johnson's research examines the ethical, social, and policy implications of technology.

Jessie Mann is an artist whose work can be found in the permanent collections of the J. Paul Getty Museum, the St. Petersburg Museum, and the Art Institute of Chicago. Mann's research is focused on the cognitive neuroscience of self-simulation and the underlying neurological mechanisms that support neurorehabilitative therapies that rely on self-simulation, such as mental imagery and avatar-assisted therapies.

Carl Mitcham is Professor of Philosophy of Technology at Renmin University of China and of Liberal Arts and International Studies at the Colorado School of Mines. His work focuses on ethics and policy issues related to science, technology, and society.

Cyrus C. M. Mody is Professor and Chair in the History of Science, Technology, and Innovation at Maastricht University. He is the author of *Instrumental Community: Probe Microscopy*

and the Path to Nanotechnology (2011) and *The Long Arm of Moore's Law: Microelectronics and American Science* (2016).

Zachary Pirtle is an engineer at NASA Headquarters and a doctoral candidate in Systems Engineering at George Washington University. Before joining NASA in 2010, he worked with the Consortium for Science, Policy, and Outcomes at Arizona State University, where he obtained degrees in mechanical engineering and philosophy, and a master's degree in civil engineering.

Joseph C. Pitt is Professor of Philosophy and of Science and Technology in Society at Virginia Tech. A Fellow of the AAAS, he is the author of four books and the editor/co-editor of thirteen others. His area of research is the impact of technological innovation on scientific change.

Nicholas Rescher is a Distinguished University Professor of Philosophy at the University of Pittsburgh, where he has also served as Chairman of the Philosophy Department and a Director of the Center for Philosophy of Science. In a productive research career extending over six decades, he has established himself as a systematic philosopher with more than one hundred books to his credit, ranging over all areas of philosophy with sixteen of them translated from English into eight other languages.

Robert Rosenberger studies the phenomenology of technology, investigating topics such as homelessness, driving impairment, classroom tech, and laboratory imaging. He is Associate Professor in the School of Public Policy at the Georgia Institute of Technology.

Evan Selinger is Professor of Philosophy at Rochester Institute of Technology. He specializes in the philosophy of technology, focusing on ethical issues of emerging technology and privacy. Currently, he is co-writing *Being Human in the 21st Century* with Brett Frischmann.

Ashley Shew, Assistant Professor in the Department of Science and Technology in Society at Virginia Tech, works on issues in philosophy of technology related to technological knowledge, animal studies, biotech, and disability studies. Her first book, *Animal Constructions and Technological Knowledge*, is under contract and forthcoming.

Tina Sikka is currently a lecturer at Newcastle University in the Media Culture and Heritage Department. She has published extensively on the importance of public participation in technological design, feminist analyses of climate science, and scientific knowledge production. She is currently writing a book on feminist empiricism, critical theory, and geoengineering that is under contract and scheduled for publication in late 2017.

Johnny Hartz Søraker, formerly a faculty member at University of Twente, works as a policy specialist for Google in Dublin, Ireland. His main research specialization lies in the intersection of information and communication technologies, philosophy, and psychology, with a particular emphasis on virtual environments and quality of life.

Julie C. Swierczek is Librarian for Primary Resources and Metadata Services at Swathmore College. She holds a master's in Philosophy from Miami University of Ohio and a Master of Science in Library and Information Science from the University of Illinois at Urbana-Champaign. Her interests include metadata, the organization of information, digital forensics, and information ethics.

David C. Tomblin is Director of the Science, Technology, and Society Scholars program at the University of Maryland, College Park. He runs a Robotics Service Learning program and an Infrastructure and Society Service Learning program. He works with a consortium

of universities, science museums, and nonprofits called Expert and Citizen Assessment of Technology to develop and do research on public engagement exercises for government agencies such as NASA, Department of Energy, the EPA, and NOAA.

Shannon Vallor is Associate Professor of Philosophy at Santa Clara University and President of the International Society for Philosophy and Technology. Her primary research expertise is the philosophy and ethics of emerging science and technology.

Yoni Van Den Eede is a postdoctoral fellow of the Research Foundation–Flanders (FWO), affiliated with the research groups Centre for Ethics and Humanism and Culture, Emancipation, Media, and Society, both at the Free University of Brussels. He conducts research into the philosophy of technology, media theory, and media ecology. He is the author of *Amor Technologiae: Marshall McLuhan as Philosopher of Technology* (2012).

Pieter E. Vermaas is Associate Professor in the Philosophy Department of Delft University of Technology, the Netherlands. He does research on design methodology, design for moral values, and validation in design research and is editing the book series *Philosophy of Engineering and Technology and Design Research Foundations*.

D. E. Wittkower is Associate Professor of Philosophy at Old Dominion University, where he teaches philosophy of technology, philosophy of social media, information ethics, and information literacy and digital culture.

Monique Wonderly is the Harold T. Shapiro Postdoctoral Research Associate in Bioethics at the Princeton University Center for Human Values. She has published in the areas of applied ethics, philosophy of emotion, and history of philosophy.

Editor Introduction

This volume provides some important things to think about when it comes to philosophical problems arising from our technologies, past, present, and future. The philosophy of technology can be seen as a relatively young field, but the questions the field poses—about the nature of humanity, technology, and our relationship to each other and the world—are ancient. We do not offer a history of the philosophy of technology here because that has already been done, and several other collections already present the classics. Rather, we offer a series of reflections on the world we live in, witness, and foresee, primarily characterized in terms of the technologies we have, emerging technologies, and technologies that may develop in the future. Many of our authors belong to a new generation of philosophers of technology and bring an excitement to their work, as adventurers and explorers of the future, laying out projects that need attention.

We have chosen ‘space’ as an organizing theme for mapping out the current and emerging state of our field. We see ourselves, our work, and our ideals as living in ethical, political, virtual, personal, inner, and outer space. There may be other spaces to explore; these categories are not intended to be exhaustive, but we think they represent the spaces that philosophy of technology is now and will be exploring in the near future. By stressing space we also stress our communal nature and the fact that we are all in this together.

The volume we are proud to bring together here features many fresh voices—and more seasoned ones that offer new material. We hope to chart the course of philosophy of technology as it will soon be written. These collected new works complement the existing canon and serve to frame the technological world in which we currently exist.

Joseph C. Pitt
Ashley Shew
Blacksburg, Virginia

Part I

Ethical Space and Experience



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Data, Technology, and Gender

Thinking About (and From) Trans Lives

Anna Lauren Hoffmann

Introduction

For scholars and students interested in topics of gender identity, data, and information technology, the current historical moment is a curious one. The proliferation of personal computing devices—from laptops to mobile phones to “smart” watches—combined with widespread internet access, means that people are producing unprecedented amounts of digital data, leading some scholars and technology evangelists to declare a “big data” revolution. At the same time, issues of sexism and gender inequality have taken on new urgency as women face increasing levels of harassment online, especially on large social networking sites like Twitter. The blame for this falls, in part, on platform owners and developers that fail to thoroughly consider role of design in promoting safety for the most vulnerable users. Finally, the emergence of high-profile transgender activists, performers, and celebrities—from Laverne Cox to Caitlyn Jenner—has brought attention to a minority population of trans, nonbinary, and gender-nonconforming people and communities that have been (until now, at least) largely overlooked, often to the detriment of the health and safety of these populations.

Of course, some would view these three trends as mostly unrelated: at a quick glance, big data, gender and sexism online, and the health and needs of transgender people seem to have little to do with one another. Against this easy assumption, however, this chapter suggests that—while not wholly reducible to one another—these three issues intersect in important ways and, in particular, they shine a light the ongoing struggles minority identities and lives face in our increasingly data-driven world. The ‘big data revolution’ cannot be divorced from the technologies and systems that support it—technologies and systems that have long struggled to account for diverse and nonnormative lives.

In the following, these three trends are woven together to further our thinking about gender, identity, and technology. The first section attends to the biases and assumptions that underwrite the idea of ‘big data.’ It argues that big data and the quantitative insights into human behavior they stand to provide are not given but, rather, they are something we make and remake in practice. The second section traces key strands of thinking about the relationship between gender and technology, offering deeper insight into the ways in which gendered biases or stereotypes are built into the practice of scientific and technological development. Finally, the third section takes these lessons and extends them to thinking about the lives and identities of gender minorities, such as transgender individuals. I should note, however, that the discussions of relevant literature throughout this chapter are not intended to be comprehensive (indeed, a fully comprehensive literature review of any section’s topic would fall outside the scope of this chapter). Rather, I mean only to hit on the most salient trends and points as they relate to and help to discuss issues of data, technology, information systems, and gender identity.

Confronting the Mythology of Big Data

The term *big data* represents many things. As Rob Kitchin (2014a) describes, the term often refers to data sets and databases that are ‘big’ along three lines: volume, velocity, and variety (the 3Vs) (67–68; see also: Zikopoulos and Eaton 2011). Under this definition, big data are unique because of their massive size (petabytes or even zettabytes), the rapidity of their production (sometimes near real time, as with data generated by social networking sites), and their diversity (they are expansive, contain data and metadata, and they can be both structured and unstructured) (Kitchin 2014b: 1). Big data are also sometimes marked by other features, like scalability (Mayer-Schönberger and Cukier 2013), the ease by which they are combined with other data (Kitchin and McArdle 2016), and their often fine-grained or detailed nature (Dodge and Kitchin 2005). Beyond technical features, big data also represent a kind of mythology. As boyd and Crawford (2012) put it, big data simultaneously are produced and thrive on a “widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy” (663).

Although the technical features of big data may raise their own practical and philosophical issues, the focus of this section is the mythology of big data. This myth—the idea that more and bigger data equals more and greater truth—is a seductive one; it suggests that the social world can be explained from a value-neutral, objective point in much the same way that the physical universe is understood through measurable and mathematically quantifiable features (Jurgenson 2014). Instead of filtering our data through the ideas and theories that make up various branches of the social sciences (like sociology, linguistics, or psychology) we can simply harness the power of today’s computers to perform automated statistical analyses on massive data sets that capture traces of human behavior. Computers can find patterns and identify correlations that humans cannot, patterns that—while not proof of causation—are basically good enough to do the job of predicting (rather than explaining) future behavior. As Geoffrey Bowker (2014) describes, such an approach seems—at least superficially—to “[avoid] funneling our findings through vapid stereotypes” (1796). Amazon, for example, deploys an online recommender system that

work[s] through correlation of purchases without passing through the vapid categories of the marketers—you don’t need to know whether someone is male or female, queer or straight, you just need to know his or her patterns of purchases and find similar clusters.
(Bowker 2014: 1796)

The seductiveness of this idea has led some big data evangelists to proclaim that we have reached the “end of theory,” a point in time where knowledge production is simply a matter of “[throwing] numbers into the biggest computing clusters the world has ever seen and [letting] statistical algorithms find patterns where science cannot” (Anderson 2008: n.p.). As Caroline Basset (2015) summarizes the idea, “Big Data ushers in new forms of expertise and promises to render various forms of human expertise increasingly necessary” through “automation of forms of data capture, information gathering, data analysis and ultimately knowledge production” (549). In Robert W. Gehl’s (2015) words, “a common refrain . . . is that we are in for a revolution, but only if we recognize the problem of too much data and accept the impartial findings of data science for the good of us all” (420). In short, big data appear to make “human expertise seem increasingly beside the point” (Basset 2015: 549).

But one can only admit the “end of theory” if one also accepts uncritically the mythology of big data. But many scholars—including those cited earlier—warn that this myth is dangerous,

as it overlooks the ways in which our very ideas about what constitutes ‘data’ are themselves framed by theoretical perspectives and assumptions. At a fundamental level, the mere act of calling some things data (and disregarding other things as ‘not data’) represents a kind of theory itself: even unstructured data rely on categories of chronological time or textual sources that have already been shaped by assumptions about the world enforced by data collection instruments. Any given data set is, by necessity, limited by its sources or its aims—no single data set, even the most massive ones, can contain all conceivable data points because not everyone or everything is conceived of as ‘data.’ Consequently, big data continue to suffer from “blind spots and problems of representativeness, precisely because [they] cannot account for those who participate in the social world in way that do not register as digital signals” (Crawford et al. 2014: 1667).

Assumptions about what constitute ‘data’ are built into the instruments and tools we use to collect, analyze, and understand the data itself. These tools “have their own inbuilt limitations and restrictions”—for example, data available through social networking sites like Twitter and Facebook are constrained by the poor archiving and search functions of those sites, making it easy for researchers to look at events or conversations in the present and immediate past but also difficult to track older events or conversations (boyd and Crawford 2012: 666). As a consequence, research conducted on or through these sites often inherits a temporal bias, and given the constraints of these social platforms, researchers prioritize immediacy over more reflective or temporally distant analyses. The mythology of big data—its appeal to automated, technologically sophisticated systems and claims to objectivity—works to obscure these biases and their limits for accounting for certain kinds of people or communities (Crawford et al. 2014: 1667). As Bowker (2014) puts it: “just because we have big (or very big, or massive) data does not mean that our databases are not theoretically structured in ways that enable certain perspectives and disable others” (1797).

To be critical scholars and students of big data we must be vigilant against a mythology. It is imperative that we pierce the veil of technological wonder and readily scrutinize big data’s claims to impartiality or neutrality and recognize that data and knowledge are made legible and valuable not in a vacuum, but in context. As Tom Boellstorff (2013) rightfully asserts: “There is a great need for theorization precisely when emerging configurations of data might seem to make concepts superfluous—to underscore that there is no Archimedean point of pure data outside conceptual worlds” (n.p.). To be sure, these limits and biases do not automatically mean that large-scale, data-intensive research is necessarily bad or unimportant. Rather, they simply underscore the continued relevance of theoretical and other types of inquiry even in the midst of a big data ‘revolution.’ As Crawford et al. (2014) argue,

the already tired binary of big data—is it good or bad?—neglects a far more complex reality that is developing. There is a multitude of different—sometimes completely opposed—disciplinary settings, techniques, and practices that still assemble (albeit uncomfortably) under the banner of big data.

(1665)

Surfacing the Role of Gender in the Design and Production of Science and Technology

The previous section challenged the seeming neutrality and objectivity of big data by reasserting the importance of paying critical attention to the social, political, and technological biases that underlie processes of collecting, analyzing, and making sense of data. This section builds on this idea by zeroing in on one particular kind of social and political bias: gender bias. It

focuses on the work of scholars and commentators that show how scientific and technological practices (and the knowledge they produce) are shaped and constrained by considerations of gender.

Early work on gender and technology focused almost exclusively on highlighting the overlooked contributions of women to the history and development of science and technology. Work in this vein sometimes focuses on women's contributions to sites conventionally associated with men—such as industry, engineering, or scientific research—and demonstrates how the narratives that have emerged around these sites have tended to privilege the work and ideas of men despite the presence and contributions of women. For example, a focus on the men who built the first electronic, all-purpose computer—the Electronic Numerical Integrator and Computer (ENIAC)—overlooks the fact that it was a team of women mathematicians that worked to program the machine and make it operational (Sydell 2014). These sorts of skewed narratives “have tended to make the very male world of invention and engineering look ‘normal,’ and thus even more exclusively male than is actually the case” (Lerman, Mohun, and Oldenziel 2003: 431). As Nathan Ensmenger (2010) summarizes, “the idea that many . . . computing professions were not only historically unusually accepting of women, but were in fact once considered ‘feminized’ occupations, seems . . . unbelievable” against a backdrop that so heavily associates computing with men and masculinity (120).

Other approaches work in a different direction, looking instead at activities and spaces historically associated with women but overlooked as significant sites of technological activity. Building on feminist critiques of Marxism that emphasized the role of unpaid and domestic labor (most often performed by women), work in this area examines the relationship between gender and technology outside of conventional sites of scientific or technological production. Cynthia Cockburn and Susan Ormrod (1993)—in their now-classic work *Gender and Technology in the Making*—examined the history and rise of the microwave oven not only in its design and development phase, but through to its dissemination into kitchens and the home. Cockburn and Ormrod (1993) show how a technology that starts out as a high-tech innovation ends up—through processes of marketing, retailing, and associations with ‘women’s work’ like household cooking—viewed as a rote household appliance, ultimately ignoring the ways that women’s specific technical knowledge (of cooking, for example) also contributed to the design, distribution, and use of a particular technology.

Despite progress in recognizing the contributions of women in the history of science and technology, however, biases still persist in our narratives about novel or innovative technologies. The story of the relatively recent and much-lauded Google Books project, for example, foregrounds the vision of Google’s founders Sergey Brin and Larry Page as well as the company’s (male-dominated) engineering teams that developed a novel way for quickly and effectively scanning, digitizing, and bringing entire library collections online (thus greatly expanding access to recorded knowledge). Lost in this narrative are the contributions of librarians (primarily women) who collected, organized, curated, and maintained the collections upon which Google Books is built (Hoffmann and Bloom, forthcoming) as well as the women and people of color who performed the manual labor of turning pages for Google’s scanning machines (Wen 2014).

Further approaches to gender and technology center not on the narratives that grow up around particular technologies, but instead on the ways in which gender biases influence the development and design of technology itself. Work in this vein seeks to uncover how sexist assumptions and stereotypes end up designed—or ‘baked’—into various systems and artifacts. For example, video games that offer only male avatars for players (or male and highly sexualized female avatars) implicitly encode the assumption that only (heterosexual) men play video games (Friedman 1996). More recently, commentators have pointed out how software

applications and personal tracking tools also fail to account for the specific needs of women. For example, the release of Apple's HealthKit for its popular mobile phones (iPhones) promised a set of comprehensive tools for tracking personal health and biometric information. However, HealthKit's first iteration failed to include a tool for tracking menstruation (Duhaime-Ross 2014). Studying the relationship between gender and technology in this way allows us to reveal and destabilize these seemingly 'natural' defaults by revealing the ways in which they actively construct biased or even harmful ideas about women. (For more thorough summaries of the state of gender and technology studies at different points in its development, see McGaw 1982; Lerman, Mohan, and Odenziel 2003; Wajcman 2009).

Finally, gender has also played an important role in normative analyses of science, helping to shed light on the moral and ethical consequences of scientific and technological progress. Sandra Harding's (1991) foundational work on feminist studies of science implored scholars to pay close attention to "the problematics, agendas, ethics, consequences, and status" of science-as-usual, that is, scientific practice and as we commonly (and uncritically) understand it (1). Doing so means going beyond simply harnessing the tools of science to explore overlooked questions or areas (like, for example, women's health needs); instead, it requires a thorough examination of the tools themselves—the methods, instruments, practices, and ethics that have come to typify scientific practice. For example, simply pointing the tools and technologies of science at issues relevant to women's lives reinforces the assumption that gender is binary and that men and women are categorically different, a problematic assumption that has historically undergirded research on sex difference (Fausto-Sterling 1985; Richardson 2013).

Against the ingrained biases and problematic assumptions of conventional scientific inquiry, many feminist researchers emphasize not one particular 'female' way of knowing, but—rather—advocate for a plurality of methods for gathering, analyzing, and making sense of the world. Regardless of method, feminist research should share—as Alison Wylie (2007) argues—at least four basic commitments: (1) research should be relevant to feminist aims of exposing and addressing questions of oppression, gendered or otherwise; (2) research should be grounded in the experiences of marginalized populations, especially women; (3) researchers should be accountable, in an ethical sense, to the persons and populations they study; and (4) researchers should be reflexive, that is, they should foreground (rather than obscure) the context and assumptions that underwrite their work. Combined, these four dimensions articulate a normative vision for science that rejects the sort of objectivity and neutrality by positivist and other understandings of science. (For a more thorough discussion of these commitments, see Crasnow et al. 2015.)

Data, Information Technology, and Transgender Lives

The preceding discussions of big data and gender, science, and technology share a careful attention to the ways in which biases, stereotypes, and problematic assumptions shape our understandings of the world. They resist any easy or uncomplicated claims to neutrality or objectivity, whether in science generally or in analyses of massive data sets specifically. For scholars critical of big data, this resistance means understanding what is admitted as 'data' in the first place (and what is left out) as well as being cognizant of the histories and politics that shape the categories we use to make sense of data. For feminist scholars of science and technology, it means bringing gender and oppression to the fore of our analyses and rejecting the supposed neutrality of scientific and technological production. Bringing these discussions together helps to open up a critical discussion of data, information technology, and the continued marginalization of gender minorities like transgender and gender nonconforming individuals.

At its simplest, the term *transgender* refers to “people who move away from the gender they were assigned at birth, people who cross over (trans-) the boundaries constructed by their culture to define and contain that gender” (Stryker 2009: 1). It is sometimes described as the opposite of *cisgender*, a term that refers to people who identify with the gender they were assigned at birth—most likely binary, either male or female (man or woman, boy or girl). Etymologically, the prefix *cis-* derives from the Latin term meaning “on this side of,” while the prefix *trans-* derives from the Latin term meaning “on the other side of.” It is important to note, however, that not all members of gender minority groups (those who are not either cisgender men or cisgender women) necessarily identify as transgender. A range of terms have emerged to describe a range of identities and lived experiences of gender—from genderqueer individuals who do not prescribe to any discrete gender category to nonbinary individuals who reject the binary categories of male/female altogether.

Going beyond this relatively straightforward definition, Megan Davidson (2007) explains that “the term *transgender* has no singular, fixed meaning but is instead . . . conceptualized by both scholars and activists as inclusive of the identities and experiences of some (or perhaps all) gender-variant, gender- or sex-changing, gender-blending, and gender-bending people” (Davidson 2007: 60). In some sense, then, ambiguity has long been integral to any conception of the term (Nataf 1996). It may, at times, include (or exclude)

transsexual people (of all operative statuses), cross-dressers, drag kings and queens, genderqueer people, gay men and lesbians who queer gender lines (such as butch lesbians), the partners of trans people, and any number of other people who transgress binary sex and gender in all sorts of named and yet unnamed ways.

(Davidson 2007: 61)

For the purposes of readability, however, this remainder of this section (following Haimson et al. 2015) uses the shorthand “trans” to refer to the transgender and broader gender nonconforming population.

Trans people are relevant to thinking about data and information systems in different ways. Contemporary practices of collecting, mining, analyzing, and otherwise making use of data represent new avenues for the exercise of social control (Andrejevic 2013). In addition, efforts to classify and categorize things are caught up in processes of power and control (Boellstorff 2013: n.p.; see also Bowker and Star 1999). As such, they represent new methods for defining and containing categories of gender—methods that may or may not account for the identities and needs of trans populations. For example, paper or online forms that offer only binary options—only male and female check boxes—pose problems for trans people trying to access health or other social services, online communities, or even dating sites. Trans women, for example, have had problems using the popular online dating application Tinder, a site that offers users only the option to identify by binary (and presumably cis) gender (i.e., man or woman). Men seeking women on the site have repeatedly reported the accounts of trans women as fraudulent based on the perceived failure of these women to meet the men’s normative standard of what a ‘real woman’ is or looks like (Vincent 2016). These reports often result in trans women’s accounts being suspended and trans users being kicked off the site.

Trans people’s struggles with information systems and biased categories also go beyond mere check boxes for gender. Many trans people, when socially transitioning genders, choose a new name for themselves—one that better reflects who they are. However, national and local policies may make it more or less difficult to legally change the name one was assigned at birth (also known as a “deadname”). As a result, trans people often find themselves being forced to

disclose information (like their deadname) through bureaucratic or administrative practices that do not account for or permit the use of chosen names that are not yet legally recognized. For example, trans people may wish to sign up to use a website like Facebook—a social networking site with more than one billion registered users—using an identity that is different from the one that appears on their birth certificate or other legal documents. However, because Facebook enforces a ‘real name’ policy, doing so is often not possible.

Beyond social networking sites, the administrative tensions generated by limited and inflexible data categories and information systems can inform all aspects of a trans individual’s life. Take, for example, the importance of name and gender information in a university context:

Because college officials use gender in assigning campus housing, determining which bathrooms and locker room students are permitted to use and deciding on which sports team students can compete, a gender marker that does not correspond to how a student identifies might mean that their institution will place them in unfair, uncomfortable, and potentially dangerous situations.

(Beemyn and Brauer 2015: 480)

More than just an administrative headache, being forced to reveal or go by the wrong gender or the wrong name can trigger feelings of dysphoria and humiliation. In some cases, it can also lead to harassment, abuse, and even death. As Dean Spade (2015) forcefully demonstrates in his book *Normal Life*, these sorts of conflicts—between prescribed categories and lived or actual identities—have severe consequences, often leading to trans people being denied housing, employment, medical or mental health care, and access to homeless or domestic violence shelters.

As discussed in the first section, the mythology of big data cannot be divorced from the systems and practices upon which the big data revolution relies—systems and practices that struggle to account for trans identities and lives. As Jeffrey Alan Johnson (2014) reiterates “it should be clear by now that, contrary to the common perception of data as an objective representation of reality, the content of data systems is an interpretation” (160). Nonetheless, making sense of data and navigating information systems, he argues, necessarily requires something like the illusion of objective representation—an illusion that “establish[es] certain state[s] of the world as within the realm of normalcy to the exclusion of others” (Johnson 2014: 162). Trans lives and identities challenge the normalized gender assumptions imposed by information systems in at least two ways: (1) categorically (through the rejection of binary gender) and (2) conceptually (through resistance to singular, fixed meanings). In doing so, they expose the limits of quantitative and big data–driven understandings of the world that rely on rigid and reductive categories in the face of fluid or shifting identities.

In addition, contemporary data science and information systems stand to further marginalize individuals (binary, trans, or otherwise) whose identities are coupled with other identities that entail other forms of oppression, such as racial or socioeconomic discrimination. Here, discussions of gender, identity, and data featured in the relatively new journal *Transgender Studies Quarterly* are instructive as they often emphasize not only gender in their analyses, but other sources of oppression—racial, ableist, classist, and beyond—as well. They embrace the idea of intersectional feminism, a concept that refers to a line of critique and activism rooted in multiracial feminist movements in the second half of the twentieth century and eventually concretized in the work of legal scholar Kimberlé Crenshaw (1989). Following Crenshaw’s (1991) powerful discussions of violence against women of color, embracing intersectional analyses means recognizing that the various sources that oppress marginalized groups—be they racism,

sexism, ableism, homophobia, classism, or beyond—cannot be understood independently of one another. Rather, they intersect in ways that generate unique experiences of oppression and marginalization. For example, “the intersection of racism and sexism factors into Black women’s lives in ways that cannot be captured wholly by looking at the race or gender dimensions of those experiences separately” (Crenshaw 1991: 1244).

Careful attention to the intersections of identity generates even richer and more diverse understandings of gender as not isolated features of identity, but as shaped and realized in complex webs of identity and social relations best understood intersectionally. Instead of an abstract category of ‘woman’ that attempts to account for otherwise disparate experiences, more complex categories emerge—suddenly, the distinct experiences of White women and Black women or cisgender women and transgender women (or Black cisgender women and Black transgender women, and so on) become available for analysis. Oftentimes, however, large-scale, data-intensive research and design fails to account for these local, context-dependent intersections and the specific experiences of violence and oppression they generate. For example, Safiya Noble (2013) has shown how the data, algorithms, and processes that produce Google search results for the term *Black girls* reduce the identities of Black women to stereotypical or hypersexualized representations only, demonstrating the unique confluence of racist and sexist oppressions faced by Black women online.

Achieving broad inclusivity in data and design is challenging even for those squarely concerned with problems of gender and technology. As Catharina Landström (2007) notes, even when “captur[ing] the ways in which technology is shaped by gender and gender is shaped by technology” (8) feminist scholars of science and technology reinforce normative standards that are constraining rather than emancipatory for some. For example, work discussed in the second section tends to “not question the definition of gender as a heterosexual coupling of opposites, female and male, masculine and feminine” (Landström 2007: 10). The conception of gender as binary (and further constrained by the heteronormative connotations of a male/female dichotomy) presents further challenges for feminist discussions of big data. As Elizabeth Losh (2015) notes in her reading of big data through the politics of the “selfie,” in certain approaches to studying big data, “gender is presented in strongly binary terms, with ‘female’ and ‘male’ as the main categories separated by a territory demarcated by a question mark” (1635). Even where online tools—and commercial models, such as Facebook’s expanded and open-ended gender identity options—make possible a diversity of representations “essential . . . for studying how under and sexuality are performed online,” defaults tend to emphasize an either/or logic (Losh 2015: 1653).

The tension between fixed, reductive categories and fluid, transitional identities is further reflected in the most progressive efforts to account for trans and other queer identities through data collection and design. In 2014, Facebook revised its gender options for users—going from two categories to fifty-eight (and eventually adding an open-ended option as well). Though this is lauded as an important move towards broad gender inclusivity, Rena Bivens (2015) has demonstrated how—at a deep level and despite the addition of expanded gender options—Facebook continues to enforce a binary logic encoded both in their business model and at the level of software. In a different example, Jack Harrison-Quintana, Jaime M. Grant, and Ignacio G. Rivera (2015) reflect on their experiences developing the National Transgender Discrimination Survey (NTDS) and note the challenges of developing “liberating versus limiting” boxes that capture identity in data in ways that do not “collapse and marginalize trans experience” but “expand and uncover the richness and complexities of trans lives” (167). Though not necessary congenial to the production of efficient and data-intensive quantitative analyses, such diverse understandings are integral to a richer and more broad ranging understanding of human behavior and experience.

Conclusion

In her feminist account of big data, Elizabeth Losh (2015) reminds us that “individuals do not float free in a loose matrix of voluntary social relations” (1651). They are, rather, constrained by power structures or practices that impose their own meanings at different levels. As boyd and Crawford (2012) put it:

Data are not generic. There is value to analyzing data abstractions, yet retaining context remains critical, particularly for certain lines of inquiry. Context is hard to interpret at scale and even harder to maintain when data are reduced to fit into a model.

(671)

The experiences of trans people extend and challenge our understandings of big data and the relationship between gender and technology in important ways. They lay bare the limits of rigid or fixed data categories for capturing fluid or multifaceted identities and they urge further examination—both theoretical and empirical—into the ways data subjects are constrained (and impacted) by biases and assumptions in scientific and technological development.

While issues of identity, data, and information systems seem to be—on one level, at least—an interesting conceptual or philosophical problem to ponder, they also expose the urgency of recognizing the very real and lived challenges these tensions and the rapid rise and adoption of data-intensive technologies and platforms generate for already vulnerable trans and queer populations. The continued exclusion from or subjugation of these populations to information systems that do not represent their lives or needs represents a continuation of the “administrative violence” described by Dean Spade (2015)—a phenomenon that we might rightly call *data violence* in order to also capture the harm inflicted on trans and gender nonconforming people not only by government-run systems, but also the information systems that permeate our everyday social lives.

References

- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. Available at: www.wired.com/2008/06/pb-theory/ [Accessed April 16, 2016].
- Andrejevic, M. (2013). *Infoglut: How Too Much Information Is Changing the Way We Think and Know*. New York: Routledge.
- Bassett, C. (2015). Plenty as a Response to Austerity? Big Data Expertise, Cultures and Communities. *European Journal of Cultural Studies*, 18(4–5), 548–563.
- Beemyn, G., and Brauer, D. (2015). Trans-Inclusive College Records Meeting the Needs of an Increasingly Diverse U.S. Student Population. *TSQ: Transgender Studies Quarterly*, 2(3), 478–487.
- Bivens, R. (2015). The Gender Binary Will Not Be Deprogrammed: Ten Years of Coding Gender on Facebook. *New Media & Society*, 1–19. doi:10.1177/1461444815621527.
- Boellstorff, T. (2013). Making Big Data, in Theory. *First Monday*, 18(10). doi:10.5210/fm.v18i10.4869
- Bowker, G. (2014). The Theory/Data Thing. *International Journal of Communication*, 8, 1795–1799.
- Bowker, G., and Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- boyd, d., and Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication and Society*, 15(5), 662–679.
- Cockburn, C., and Ormrod, S. (1993). *Gender and Technology in the Making*. Thousand Oaks, CA: Sage.
- Crasnow, S., Wylie, A., Bauchspies, W. K., and Potter, E. (2015). In Edward N. Zalta (ed.) *Feminist Perspectives on Science*. In *The Stanford Encyclopedia of Philosophy*. Stanford, CA: Metaphysics Research Lab, Stanford University.

- Crawford, K., Miltner, K., and Gray, M. L. (2014). Critiquing Big Data: Politics, Ethics, Epistemologies. *International Journal of Communication*, 8, 1663–1672.
- Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, 1989(1), 139–167.
- Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color. *Stanford Law Review*, 43(6), 1241–1299.
- Davidson, M. (2007). Seeking Refuge Under the Umbrella: Inclusion, Exclusion, and Organizing Within the Category Transgender. *Sexuality Research & Social Policy: Journal of NSRC*, 4(4), 60–80.
- Dodge, M., and Kitchin, R. (2005). Codes of Life: Identification Codes and the Machine-Readable World. *Environment and Planning D: Society and Space*, 23(6), 851–881.
- Duhaime-Ross, A. (2014). Apple Promised an Expansive Health App, So Why Can't I Track Menstruation? *The Verge*. Available at: www.theverge.com/2014/9/25/6844021/apple-promised-an-expansive-health-app-so-why-cant-i-track [Accessed April 16, 2016].
- Ensmenger, N. (2010). Making Programming Masculine. In T. J. Misa (ed.) *Gender Codes: Why Women Are Leaving Computing*. Hoboken, NJ: Wiley, 115–142.
- Fausto-Sterling, A. (1985). *Sexing the Body: Gender Politics and the Construction of Sexuality*. New York: Basic Books.
- Friedman, B. (1996). Value-Sensitive Design. *Interactions*, 3(6), 16–23.
- Gehl, R. W. (2015). Sharing, Knowledge Management and Big Data: A Partial Genealogy of the Data Scientist. *European Journal of Cultural Studies*, 18(4–5), 413–428.
- Haimson, O. L., Brubaker, J. R., Dombrowski, L., and Hayes, G. R. (2015). Disclosure, Stress, and Support During Gender Transition on Facebook. In *CSCW '15*. Vancouver, BC, Canada: ACM, 1176–1190.
- Harding, S. (1991). *Whose Science? Whose Knowledge? Thinking From Women's Lives*. Ithaca, NY: Cornell University Press.
- Harrison-Quintana, J., Grant, J. M., and Rivera, I. G. (2015). Boxes of Our Own Creation: A Trans Data Collection Wo/Manifesto. *TSQ: Transgender Studies Quarterly*, 2(1), 166–174.
- Hoffmann, A. L., and Bloom, R. (forthcoming). Digitizing Books, Obscuring Women's Work: Google Books, Librarians, and Ideologies of Access. *Ada: A Journal of Gender, New Media, and Technology*, 9.
- Johnson, J. A. (2014). From Open Data to Information Justice. *Ethics and Information Technology*, 16(4), 263–274.
- Jurgenson, N. (2014). View From Nowhere. *The New Inquiry*. Available at: <http://thenewinquiry.com/essays/view-from-nowhere/> [Accessed April 16, 2016].
- Kitchin, R. (2014a). Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, 1(1), 1–12.
- Kitchin, R. (2014b). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequence*. Thousand Oaks, CA: Sage.
- Kitchin, R., and McArdle, G. (2016). What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets. *Big Data & Society*, 3(1), 1–10.
- Landström, C. (2007). Queering Feminist Technology Studies. *Feminist Theory*, 8(1), 7–26.
- Lerman, N. E., Mohun, A. P., and Oldenziel, R. (2003). The Shoulders We Stand On/The View From Here: Historiography and Directions for Research. In N. E. Lerman, R. Oldenziel, and A. P. Mohun (eds.) *Gender & Technology: A Reader*. Baltimore, MD: The Johns Hopkins University Press, 425–450.
- Losh, E. (2015). Feminism Reads Big Data: “Social Physics,” Atomism, and Selfiecity. *International Journal of Communication*, 9, 1647–1659.
- Mayer-Schönberger, V., and Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray.
- McGaw, J. A. (1982). Women and the History of American Technology. *Signs*, 7, 798–828.
- Nataf, Z. I. (1996). *Lesbians Talk: Transgender*. New York: Scarlet Press.
- Noble, S. (2013). Google Search: Hyper-Visibility as a Means of Rendering Black Women and Girls Invisible. *InVisible Culture*, 19. Available at: <http://ivc.lib.rochester.edu/google-search-hyper-visibility-as-a-means-of-rendering-black-women-and-girls-invisible/>.
- Richardson, S. (2013). *Sex Itself: The Search for Male and Female in the Human Genome*. Chicago: University of Chicago Press.

- Spade, D. (2015). *Normal Life: Administrative Violence, Critical Trans Politics, and the Limits of Law*. Durham, NC: Duke University Press.
- Stryker, S. (2009). *Transgender History*. Berkeley, CA: Seal Press.
- Sydell, L. (2014). The Forgotten Female Programmers Who Created Modern Tech. *NPR: All Tech Considered*. Available at: www.npr.org/sections/alltechconsidered/2014/10/06/345799830/the-forgotten-female-programmers-who-created-modern-tech [Accessed April 16, 2016].
- Vincent, A. R. (2016). Does Tinder Have a Transphobia Problem? *The Huffington Post*. Available at: www.huffingtonpost.com/addison-rose-vincent/does-tinder-have-a-transp_b_9528554.html [Accessed April 16, 2016].
- Wajcman, J. (2009). Feminist Theories of Technology. *Cambridge Journal of Economics*, 1–10.
- Wen, S. (2014). The Ladies Vanish. *The New Inquiry*. Available at: <http://thenewinquiry.com/essays/the-ladies-vanish/> [Accessed April 16, 2016]. doi:10.1093/cje/ben057.
- Wylie, A. (2007). The Feminism Question in Science: What Does It Mean to “Do Social Science as a Feminist”? In S. Hesse-Biber (ed.) *Handbook of Feminist Research*. Thousand Oaks, CA: Sage.
- Zikopoulos, P., and Eaton, C. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. New York: McGraw Hill.

Discrimination

D. E. Wittkower

Langdon Winner's famous article, "Do Artifacts Have Politics?" (1980), must be the first thing mentioned in any discussion of what philosophy of technology has contributed to our understanding of discrimination. The examples addressed, most of all the famous 'racist bridges' of Robert Moses—allegedly¹ built low in order to specifically exclude New York City buses, and the kind of person more likely to be using public transportation, from certain beaches—make clear that artifacts can be said at least to have political effects, including advancing racial discrimination.

Winner's work has been highly cited, and rightly so, but it is not a full theory of *discriminatory technologies*, and it is not grounded in multiple theoretical perspectives in order to maximize its applicability within the field. In the following, I will develop such a theory; connect it with Heideggerian, Latourian, and postphenomenological theoretical structures; and demonstrate its applicability to a wide and widening range of forms of normativity, exclusion, and discrimination. This analysis will be limited to an American cultural context, as cultural constructions of discriminatory norms, especially those of race, gender, and religion, are far too varied across regions and societies to be meaningfully addressed simultaneously, and my goal in this article is in-depth analysis rather than cross-cultural exploration. I hope that readers in Germany will be able to see parallel issues presented to Turks; that Israeli readers will see parallel problems in a different religious social normativity; that those from Brazil and India will see similarities in the way that their societies normativize 'whiteness' within more of a spectrum, but still resulting in significant discrimination; that Canadians and Swedes will see connections to the cultural erasure of their indigenous peoples; and so on.

First, it must be asked, what it would be to have a full theory of discriminatory technologies? Next, it must be asked, what we are to make of the idea of a 'discriminatory technology'? Following this, we may approach the three theoretical groundings described earlier in order to provide support to the theory and to provide direction in seeing what kinds of artifacts it can help to identify and understand as technologies of discrimination.

What Would Be a Full Theory of Discriminatory Technologies? The Ontology of a Band-Aid

Using a brilliant illustration from Preston Wilcox's *White Is*, Richard Dyer (1997: 41) prompts us to consider the Band-Aid as a paradigmatic example of normative whiteness. His work in this famous book, *White*, seems to me among the finest examples of the field of cultural studies one could find. What philosophy of technology can add to this is a movement from reading the artifact as a text to looking at the way the object is concretely active in the construction of exclusionary normativity.

Band-Aids come in a variety of shapes and sizes, showing their responsiveness to a variety of contexts of use. In philosopher Luciano Floridi's language (2014), it needs to have the right protocol to fit its prompter—in this case, the minor cut in the skin. This is why it is made to minimize infection, with a mesh to discourage adhesion to the healing flesh, and available in different sizes in order match up right with the naturally occurring diversity of bleeding gashes.

There are limits to the diversity of adhesive bandage sizes and shapes, however. Three sizes to a pack is good enough to cover most cases satisfactorily well, and we recognize it would be unreasonable to expect just the right bandage for each particular wound. Having only a single size is not responsive enough to the relevant cases, while having a dozen different sizes in every box is far more responsiveness than is necessary and would likely result in a bunch of odd shapes and sizes—which would pile up in half-empty boxes accumulating in the corners of our medicine cabinets, eventually to be discarded.

The invariance of the color of adhesive bandages, until relatively recently, places significant variance from 'White' skin in this same category of irrelevance. Dark skin is apparently a prompter to which it is not necessary to design a protocol to respond. This may be an effect of 'color-blindness': the (White) product designers failed to consider that 'flesh colored' might not be the same thing for everyone. Although exceedingly unlikely in this particular case, this could be an effect of conscious discrimination akin to Moses's bridges, where the designers specifically chose to design a Whites-only product. This could simply reflect the reality of the market, where a color is chosen that will match best for the largest set of similarly colored consumers. In any case, the ontology established by the object is the same: the function of the bandage's color is to match the skin; when it fails to do so, it implicitly claims that *this* flesh is not 'flesh colored.' The proper function of the technology contains within it an ontology that may define some persons as normative and others as lesser or deviant Others.

A full theory of technologies of discrimination should engage with technologies at this level—not by reading them as texts, not by producing analyses of particular effects or even kinds of effects of technology, but by theorizing how those technologies embody, transmit, and produce ontologies of normativity that result in privilege and discrimination.

What Is a Discriminatory Technology?

Without saying anything too contentious about a proper definition of 'technology,' we can perhaps say that a common-sense description might be that a technology is a way to get something done. By speaking of a 'discriminatory technology' we must mean a way of getting something done which produces a discriminatory effect.

By speaking of 'discrimination' we clearly do not intend the word in the sense in which a gourmand has 'discriminating taste,' although the two senses are related. Discrimination in the political sense has to mean something like *when a morally irrelevant characteristic is allowed undue influence in a determination of individual or distributive justice*. For discrimination of this political kind to take place, there must first be discrimination in the amoral sense of drawing distinctions—in this case, distinctions among persons.

This may seem to be a trivial conceptual point: of course a distinction must be first made before it can be given undue influence. But this point is of deep and historical consequence, for, arguably at least, the most prominent way that discrimination has been overcome is not by equalizing the judgments made about one group who has been distinguished from another, but instead by ceasing to make the distinction between these groups to begin with.

We see this in the history of the meaning of 'White' in a racial sense. In the colonial period, we find records of Blacks and Irish rising up against Whites, although today we consider the

Irish to be White. The Polish have similarly and recently disappeared as a category distinguished from 'Whites.' Hungarians and Bulgarians, in the mid-twentieth century, were subject to racial slurs (*bohunk*, *hunky*) largely unrecognizable today to the people that these words were meant to Other and to denigrate.

Religious and language differences play at least as much of a difference as more distinctively race-related in determining who counts as White. In European history, Spaniards and Italians have not always been considered White, especially those of Muslim faith. As Dyer points out (1997: 42), Jews seem to have been considered Black for some time, and became only White in the latter part of the twentieth century. Semitic persons, especially when Muslim or immigrants, are Caucasian but nonetheless are often not White—the same is true of Latina/os, especially when English is a second language.

What then is it to be White? The approximate answer from critical race theory is that being 'White' just means that it is not noted in any consequential way that you are raced at all. If you are encountered in the context of a racial identification, you are a person of color (PoC); if you are not, you are White. In this way we see that identification as White is a lack of judgment rather than a concrete claim: that you are White means nothing besides that you have not been identified as something else.

This judgment requires training—just as does having a discriminating palate. The way in which formerly non-White persons become White involves a decrease in the weight placed upon prejudicial claims against the minority in question, but it also involves a decrease in the amount of training people receive in identifying those persons as a group in the first place. The Nazis produced materials specifically designed to help Whites to identify and out Jews, but much more innocuous racial caricatures, as in political cartoons, play a similar role.

This is why status as White or PoC is a matter more of how we are interpreted rather than a matter of fact, although matters of fact form the basis of any interpretation, and certainly can limit the range of interpretation available. Someone of mixed race may pass as White and identify as White, and have no idea that they have non-White ancestors. A person of European descent with curly hair may be darker skinned than a person of African descent with straight hair, but we have been trained to interpret racial cues in such consistent and nuanced ways that there may be little or no controversy about the whiteness of the former and the blackness of the latter. Many people, including myself, are raced differently in different contexts or when wearing clothes or hairstyles with or without ethnic markers. Few people today are raised in environments in which they are trained to recognize my features or my surname as racial markers—but some are. To most, I am White, but I have been threatened with violence by White racists for my non-White identity.

This is what we mean when we say that race or gender are socially constructed: they are the product of human labor, manufactured using some physical basis, genetic and phenotypic, but reducible to or determined by that basis in only the same kind of limited sense that other manufactured goods are reducible to or determined by their raw materials.

An invisible starting point in our encounter with one another, prior to the construction of difference, was described by Martin Heidegger (1927/1996) as *das Man* ("the One," or "the they"). By "the One" he means only to indicate the approach to others named by the word *one* in phrases like "one doesn't do that." The "One" who does this or doesn't do that is no person in particular, or even a description of a variety or totality of actually existing persons, but is instead a set of expectations we are trained to have, on the basis of which we judge others and ourselves. The One is normativity of all kinds: One is kind, One doesn't tell lies, and One sets the table with the fork on the left of the plate and the knife on the right. And one is called to account when one doesn't do what One does!

Heidegger describes the normativity of the one in terms of an “average everydayness” that requires a “leveling down” on our parts (1927/1996: §27). It is, in his account, easy enough for us to fall into the One—to live our lives as One does and to believe what One believes and remain free from having to come up with our own values, judgments, beliefs, etc. Our care and anxiety before our own death may shake us loose from our fallenness into and entanglement with the One, because what One does and what One believes offers us no consolation before or explanation of the necessity of our own death. The need to come to terms with our own death forces us to try to make sense of our lives, and that, in turn, forces us to try actually to develop some idea of what we should do with our lives and why any of it matters. This is what Heidegger means by authenticity: to figure out for ourselves what to do and why (or die trying) rather than remaining lost in the One.

Heidegger fails to recognize that the One is constructed in ways that are directly exclusionary of many or most people, making the worry about falling into the One and failing to confront one’s individuality much more relevant to straight, White, cisgendered, affluent, Christian men than to the rest of us, who are reminded more or less regularly of our nonconformity with the One. W.E.B. Du Bois expressed this phenomenologically in the famous opening of *The Souls of Black Folks*:

[T]he negro is a sort of seventh son, born with a veil, and gifted with a second sight in this American world,—a world which yields him no true self-consciousness, but only lets him see himself through the revelation of the other world. It is a peculiar sensation, this double consciousness, this sense of always looking at one’s self through the eyes of others, of measuring one’s soul by the tape of a world that looks on him in amused contempt and pity. One ever feels his two-ness,—an American, a Negro; two souls, two thoughts, two reconciled strivings; two warring ideals in one dark body, whose dogged strength alone keeps in from being torn asunder.

(1903/1994: 2)

Why is double consciousness part of the Black experience, but not part of the White experience? It is tempting to say that the White experience is relevant for everybody to consider because of the wealth, power, and majority status of Whites, but this is incoherent. There is no ‘White experience,’ just as there is no ‘whiteness,’ only a lack of being raced. We must say instead that those who are regularly and predictably raced must consider themselves both as persons (as One) and as persons who One identifies as Other—as subject to normativity, but also as always already deviating from norms. One is not Black; One is not a woman; One is not gay; One is not transgender; One is not poor; etc.

Now, unlike “One should pursue a career” or “One ought to get married,” no one would ever actually say or even think “One isn’t Black” or “One isn’t a woman,” but the ‘average everydayness’ of White/male/straight/etc. normativity can be seen at work quickly enough in countless ordinary examples. The German term *das Man* is clearly enough gendered, and parallels long-standing English cognate usages, such as using *man* or *mankind* to refer to all humans. Here, of course, we can’t say that the view of male as ‘normal’ and female as ‘different’ can be accounted for by majority status, since women comprise approximately the same proportion of the species as do men. Regarding sexual orientation, we might consider the question, “when did you choose to be straight?” This question, of course, is never asked because our images and stories establish heterosexuality as a default condition. The effect is perhaps most striking when the population matching the One is clearly a minority, as in the body shapes and sizes that are treated as representative of femininity and masculinity in our fashion magazines and films.

As the hammer disappears in our experience of the work, integrated with its system of objects to which its affordances are tailored, so too does a normative embodiment appear to us as ready-to-hand. Only when the hammer is broken, in Heidegger's famous analogy (1927/1996: §15–16), does the artifact become present-at-hand as an object to be perceived. So long as the hammer is working, we have questions only about our projects and purposes along with it; once it is broken, we need to ask how it's supposed to work, and why it had been put together this way rather than some other way.

The normativity of the One is found in the obtrusiveness of persons rather than artifacts when there is a mismatch—to find oneself as Other rather than One is to have this mismatch attributed to one's own brokenness rather than the object's. The native English speaker, in countries where One speaks English, can wonder why One has to “Press 1 for English.” The ‘overweight’ airplane passenger, however, is more likely to blame themselves rather than the seat when it cannot accommodate their bulk. To return, now, to the Band-Aid, it is because One is White that the dark-skinned user doesn't simply say “They made this the wrong color,” but feels instead that they are the wrong color. The presence-at-hand of the artifact indicates a brokenness, a mismatch between the object and its application; the normativity of the One informs us of on which side of the relation the fault lies.

With this, we can now offer a definition of *privilege*: Privilege is the invisibility of our attributes caused by their fallenness into the One, which invisibility prevents these attributes from being perceived as meaningful. To have privilege means that the starting assumption, when something goes wrong, is that it is the world rather than the self that is broken; to lack privilege means that this, at a minimum, is not assumed.

This is why Peggy McIntosh's (1989) influential account of privilege as tools in an “invisible knapsack” is often off-putting to those who are most privileged: the tools of privilege disappear entirely to those for whom they are ready-to-hand. These aspects of privileged persons are not well-designed hammers or properly colored adhesive bandages, but are so integrated with the One that they do not appear as tools at all. For this reason and within this Heideggerian theoretical context it is more useful to think of privileges not as tools for getting things done, but instead as the failure of our attributes to appear at all; as our predetermined belonging to the One.

Now, to bring this together, we have seen that a precondition to discrimination is the drawing of distinctions. Privilege is when these distinctions do not appear, and the invisibility of whiteness and masculinity and Christian faith and so on is produced by their averageness, their fallenness into the One. Where these distinctions do appear, when they are present-at-hand rather than ready-to-hand. They are subject to being treated as having explanatory power, leading to discrimination in a political sense.

Let us now answer the question: What is a discriminatory technology? A discriminatory technology, so long as we understand technology in its broadest sense as a way of getting something done, is a method by which we background some set of attributes and foreground others, causing some attributes to disappear and become transparent and others, by contrast, to stand out. These technologies—or, techniques, if you prefer a narrower definition of ‘technology’—may include language, images, stories, policies, and objects.

Philosophy of technology is concerned with techniques, or technology in its broadest sense—but it is also and especially concerned with artifacts, or technology in a narrower sense. Having established a general theory, we can now go on to look at discriminatory artifacts.

Discriminatory Artifacts

We do well to listen to the knowledge of objects. The subtleties of the traditional size and shape of the hammer's handle contain within it a knowledge of how best to hold it, and weights it well

so that the artifact disappears into the user's experience of control and precision, even when that shape is transmitted through manufacturers who merely imitate its form while unknowing of its function. But objects can contain and transmit prejudice as well as knowledge, and the hammer—which, through its generations of traditional manufacturing, has come to conform itself to a man's hand—may, just as unknowingly, bring a feeling of awkwardness rather than authority to users with smaller hands and lithe fingers.

This occurs not only through artifacts that have discriminatory outcomes, as in Winner's case of Moses's racist bridges, but through artifacts that establish exclusionary norms. Technologies can embody discriminatory presumptions through a Latourian delegation (1992, 1999), where social values are enforced through material implication, surviving through replication of design long after their designers unthinkingly built their discriminatory values into the objects. Carpenters' tools were built for men's hands because men were carpenters. Those tools, though, have become discriminatory because our tools, unlike us, continue to act as if they believe that One does not build houses if one is a woman; they make the job easier for men than women, rendering to (most, larger-handed) male carpenters an unearned privilege and to (most, smaller-handed) female carpenters an undeserved disadvantage.

Embodiment Technics

This represents a discriminatory effect delegated to artifacts in an embodiment relation to the world, in Don Ihde's sense (1990). Embodiment technologies allow the user to access or affect the world by withdrawing into the user's experience of self. Perhaps the clearest example is corrective lenses: a good pair of prescription glasses should reveal the world while themselves disappearing from our experience; a pair of glasses that we notice constantly isn't a good pair of glasses for us. Another example is clothing, which should allow us to experience our environment in a pleasant manner, making the experience of the clothing disappear. Ill-fitting shoes or clothes that are too heavy or too light for the weather fail to disappear from our experience in the way they should.

This withdrawal from experience produces an 'enigma position' between the user and the technology. Because users experience the world as part of a human-technology hybrid, when something goes wrong, user must decouple themselves from the technology in order to ascertain where the breakdown has occurred. When our vision is blurry, we take our glasses off and clean them; when the blurriness remains, we remove our glasses again, but try rubbing our eyes instead.

The hammer that fails to withdraw into the practice of carpentry indicates a problem. It feels too heavy, or badly weighted. It is awkward. But if it is a standard design, and seems to be a good design for One to work with, we are likely to think that the fault lies with us rather than the object. And so the woman using such a tool is more likely to think that she's not good at what she's doing rather than that the tool isn't good for her.

The adhesive bandage is another example of discriminatory object in an embodiment relation: it is designed to be the color of One's skin, but fails to withdraw from the dark-skinned user's perception.

Strollers enforce regressive gender norms through a breakdown in embodiment relations as well. They are made for their average (female) user, being uncomfortable and obtrusive to most men and to tall women. In my experience, I found that not only were stroller handles too low for me to use comfortably, but the brakes on the rear wheels of our stroller were placed so that I had to train myself to take smallish steps—my natural stride made me step on the brakes by mistake whenever I started pushing the stroller at a moderate pace.

Kitchen counters are similarly designed for the average woman's height. By placing these average gender differences into design, these activities are made more difficult for men, and so

tend towards remaining women's work. Our attitudes change with the times, but the attitude written into our objects continue to make regressive claims about what One ought and ought not to be doing.

Imaging technologies provide a series of illuminating examples of discriminatory embodiment technologies. In the realm of three-dimensional imaging, new media theorist danah boyd (2014) has pointed out that constructing an immersive virtual environment through motion parallax depth cues works well for most men, but poorly for most women, because of differences in how most women's and most men's brains process visual information. The reliance of most virtual reality systems on motion parallax systems thus produces a sexist effect, where most men are able to be virtually embodied in virtual worlds (through, e.g., *Oculus Rift*), but most women are excluded, unable to allow their bodies to disappear into virtual embodiment due to the nausea that motion parallax imaging often produces in their physical bodies.

Lorna Roth (2009) has written an excellent media history of the Shirley Card—a standard photograph used to calibrate photographic printing equipment. Until the 1990s, the standard Shirley Card depicted a White woman, and photo labs calibrated to print her face with good clarity and contrast, leading to generations of poor photographic representations of dark-skinned persons, where Black faces and bodies sometimes appeared as shadows or blots with few distinguishing features other than White teeth and eyes—not just in the media, but in family snapshots. Compounding this is that even the film formulations were directed toward accurate reproduction of White faces, to the extent that Polaroid produced a specific camera for the apartheid South African government in order to produce identifiable ID card photographs of Black citizens, featuring a “boost button” that would increase the lumens of the flash by 42%; the average increased amount of light absorption of Black African skin (Smith 2013).

Roth describes this as “dysconscious racism” through a “technological unconscious” in which “a global assumption of ‘Whiteness’ [was] embedded within [photographers] architectures and expected ensemble[s] of practices” (2009: 117). A substantial change in producing less-discriminatory photographic formulation was the 1995 commercial release of Kodak Gold Max, which was referred to as being able “to photograph the details of a dark horse in low light” (Richard Wien, Executive, Kodak, Rochester, NY, personal communication, August 18, 1995, quoted in Roth 2009: 121–122).

In these cases, the embodiment relation of the photograph—where the photograph allows us to see an image as if we had been there in the place of the camera—contains a racist overlay delegated into its technological unconscious, making Black people appear darker and more indistinguishable or interchangeable, which is a racist way of seeing strongly resonant of the conscious racist depictions of Black people through minstrel shows and the Little Black Sambo and Mammy stereotypes. Through a technological version of DuBois's double consciousness, this racist representation was even reflected back to dark-skinned persons, with family and self-portraits becoming a distorted funhouse mirror in which they saw themselves not as they would appear when stretched tall or squashed short, but as they would appear to a White racist gaze. It is thus unsurprising that many PoCs view selfies in an age of Instagram and digital filters as liberatory and empowering in a way that may not be apparent to Whites. Our technological assemblages have long been oriented in order to reflect Whites' faces back to them as they see themselves, but the new level of control and authorship in digital photography has allowed PoCs to appear before any viewer of their selfie in accord with their own self-perception. Prior photographic technologies made the One's gaze racist; the digital selfie can allow One to gaze upon the PoC's selfie without the racist filters of the past.

But the One formed through delegation into the technological unconsciousness still embodies a racist gaze in other ways. Facial recognition software is still sometimes, perhaps often,

programmed to work according to contrast recognition algorithms that are calibrated to White faces and fail to recognize Blacks as persons, as demonstrated in one viral video entitled “HP computers are racist” (2009). In classroom discussion, a student shared a similar experience with me: he worked in a grocery store where employees clocked in and out using fingerprint recognition scanners that, he said, regularly failed to recognize his Black co-workers. In another similar algorithmic misrecognition, Google Photo’s image recognition software—presumably not properly primed with enough images of PoCs and not programmed properly to recognize humanity through universal human rather than White human attributes—automatically tagged Black people as gorillas (Mullen 2015). The effect of the technology is to construct Black persons as nonpersons: in some sense it is obviously right, despite the anthropomorphism of the claim, to describe such technologies as racist, for they fail even to notice the existence or the humanity of some persons on the basis of skin color.

To avoid creating discriminatory technology, it is insufficient to ensure that we do not build in racist/sexist/etc. values, for the One that we design for places statistical regularities into the architectures through which we act in the world, and so enforces those regularities upon those minorities to whom they do not apply. To have a technological unconscious that does not exhibit bias, we must take affirmative action to include minority bodies in our design spaces so that embodiment technologies fit with the diversity of user bodies. Nondiscriminatory technologies require not only designers who are not bigots, but also designers who are actively anti-racist, anti-sexist, anti-ableist, anti-transexclusionary, anti-heterosexist, and so on, because the averageness of the One will always be exclusionary of difference.

While this in no way exhausts or systematically covers ways in which embodiment technologies can be discriminatory, the most I can hope to accomplish in a chapter of this length—or even of ten times its length—is to address patterns and kinds of discriminatory artifacts. Ihde’s division of kinds of human-technics relation provides a useful structure in ensuring that I cover different sorts of such artifacts, and so, incomplete as my coverage of embodiment technologies must be, I will move forward into what he addresses as “hermeneutic technics” (1990: 80).

Hermeneutic Technics

Hermeneutic technics represent a part of a world to a user in such a way that the world is represented by rather than seen through the technology. Ihde’s phenomenological variation regarding the thermometer provides a useful contrast between embodiment and hermeneutic technics (1990: 84–85). An embodiment-based thermometer can be imagined: a metal strip going through a wall that would transfer heat efficiently, so that one’s hand could be placed on it in order to feel temperature, hot or cold, as if one’s hand were outside rather than inside. The thermometer as we know it, though, does not transfer an experience to the user, but instead translates information about the world into a different nonisomorphic format, where it can be ‘read.’ The ‘what-it’s-like’ of 30 degrees is phenomenally nonobvious and must be learned by rote, just like the ‘treeness’ of the written word *tree*—and, indeed, Ihde identifies written language as another hermeneutic technology.

In hermeneutic technologies, it is not the technology that disappears into the user’s experience of the world, but rather the world that disappears into the user’s experience of the technology, forming an enigma position between the technology and the world rather than one between the user and the technology. When the “check engine” light comes on, for example, we know that there is either something wrong with the engine or something wrong with the “check engine” indicator itself, but we don’t know which until the engine is decoupled from its representation and each is investigated separately.

Hermeneutic artifacts can become discriminatory when they fail to properly represent a person, or when they represent a nonhuman part of the world in a way that systematically excludes minority ways of understanding the world.

In the former case, we can think simply of the common understanding of the words *woman* and *man* as attached to biological sex markers, such as genetics. Insofar as we feel the need to identify a transgender woman, for example, as a trans woman, we assert that if One is a woman, one is cisgendered, and thus that a trans woman is both woman and not-woman, and another form of dual consciousness is constructed. This is made clear by the need for the term *cis woman*: if *woman* is the general term, and *trans woman* is taken to be a subset of women, then we need another word for the remainder of the set of women—otherwise transgendered women are conceptualized as not ‘really’ women, for they are not in the subset of women₁ (in the general sense) who are women₂ (in the sense of not being trans women).

The seeming objectivity of technical artifacts creates the normativity that makes this failure to capture difference discriminatory. Consider for example the résumé as a hermeneutic technology. A résumé represents a person through a well-defined set of filters, through which much of one’s life is excluded in order to provide efficient evaluation of candidates based (ideally) on relevant characteristics only. But what the same qualifications and years and kinds of experience represent in lives as they are lived can vary widely. Gaps in employment are often viewed unfavorably, as they are often taken to mean that a candidate has difficulty holding down a job. These gaps may indicate that a potential employee is unreliable, but they may instead indicate that the applicant left the labor pool for some time in order to raise or even simply to bear children—and this more favorable possibility applies more often to female rather than male applicants. Similarly, much stock may be placed in the reputation and name recognition of an applicant’s degree-granting institution, disadvantaging applicants who attended historically Black colleges and universities (HBCUs) that may be less well-known to White managers than similarly distinguished and high-quality non-HBCU institutions. In these ways and many others, the seeming-objectivity of the hermeneutic artifact—in this case, the résumé—conceals the White/male/cis/ableist normativity of the applicant whom it is designed to objectively represent.

Hermeneutic artifacts may also interpret parts of the world other than persons in a way that has discriminatory effects. The standard school and business calendars are obviously discriminatory in the way that they are organized around Christian holidays in majority-Christian-faith contexts. While it is a matter of practicality to plan around those days that most employees, managers, teachers, and students will wish to have off, the effect is to construct a set of holidays that One celebrates. Calendars are then invisible to Christians and opaque to those of other faiths, with barriers constantly and annually placed in between their organizationally provided schedules and the schedule of their faith practices. Non-Christians of faith must out themselves and request allowances to live their life in accord with their interpretation of the world as a place of religious observance; Christians need only decide whether to practice or not, for their world is largely organized to provide the opportunity. It is striking that ‘Spring Break’ for many public schools in the secular United States is moved year to year to correspond with Easter.

The discriminatory nature of the artifact of the calendar is clearest with Christmas, where it takes on additional complications. When non-Christians reject the dual consciousness of being expected to celebrate holidays that they do not celebrate—objecting to ‘Christmas parties’ at work and not smiling and saying “you too” when wished a “Merry Christmas”—this is described as an attack in the ‘War on Christmas.’ Amazingly, there is an entire genre of ‘entertainment’ dedicated to naming and shaming as ‘Grinches’ and ‘Scrooges’ those who fail to be appropriately enthusiastic about a religious holiday that One celebrates, but which a very large and diverse portion of the population of secular nations do not celebrate. At the same time as

pressure is exerted on all persons to participate in Christmas as a universal holiday, religious minorities' inclusion is disallowed by the fight to 'keep the Christ in Christmas'—to maintain its character as a religious holiday, even as those who are not Christians are expected to celebrate it.

Hermeneutic technics construct reality as One lives it as reality itself, leaving many to live in the world of their own experience with an overlay of life as lived by the One from which they are excluded, a part of a social reality that consistently fails to recognize and reflect their world.

Alterity Relations

Ihde's category of alterity relations (1990) with technology are of a different kind than embodiment and hermeneutic technics: here, the artifact does not mediate a relation between a user and the world, but instead the user interacts with the technology as such, and the relation of the technology to the world is not of central importance to that interaction. To some extent, any media representation of persons is an alterity relation: for example, consider the disproportionate representation of minorities as criminals in television programming. We don't take this representation to be representative, and its function is entertainment rather than truthful and accurate representation of reality. There is, however, an obvious spillover effect: it would be wrong to criticize any particular story for choosing a PoC as antagonist, since it is just a story and makes no reality claim, and yet a society in which Black people are seen regularly on the screen with weapons is a society in which police officers are more likely to think, in the heat of the moment, that the thing a Black citizen is holding is a gun rather than a cell phone.

These cases having to do with media representation are, however, already well covered in other fields, and particular media objects don't clearly fit a common-sense definition of technology as a way of getting something done. A crossover case might be the rightly controversial sometime use of mugshots for police target practice, which reinforces racial disparities in assumption of criminality and use of deadly force against minorities already subject to disproportional police attention and aggression. But the widest range of alterity technics that affect how we think of ourselves and others is in gaming, where we interact with virtual persons in complex and rich ways.

In games, female characters often have exaggerated secondary sex characteristics, are otherwise slim, and wear revealing and impractical outfits. But the form of interaction that gamers have with and through them establishes a deeper set of troubling norms as well. Female characters are often in support rather than lead roles, cast as magic users, healers, rogues, and archers. Exceptions abound, but women are more often heroes insofar as they allow the predominantly male characters to succeed, and do not as often succeed on their own or with support from male characters. When male characters act as support, it is often as "tanks" whose physical toughness allows for frail "glass cannon" characters to survive long enough to be effective. Women, then, who choose to play female characters must often enact the idea that women are secondary to men, whose excellence comes from their ability to allow men to triumph—and women who play primary roles, being supported rather than supporting, often do so in drag through male avatars. Male gamers, of course, play as female avatars as well, and role-play reinforces the same message to them: being a woman usually means being a helper, and being a helper usually means being a woman.

It's worth emphasizing that the point is not that this is how these characters are depicted, but how they are enacted. Female support characters may be strong and confident and have their own stories, and male lead characters may be muscled and scantily clad cardboard cutouts, but the mechanics of play often contain a clear script of male primacy, and success in the game requires adopting and acting out this view of gender roles.

This constitutes an alterity relation insofar as we are interacting with avatars on their own terms. These dynamics are present whether or not male lead and female support characters are played by male or female gamers, and very often the gender identity of other players is not known, or the game is played with the gamer controlling a team of variously gendered avatars on her own. Although here we interact with avatars through an alterity relation, gaming also represents a kind of embodiment relation insofar as we enact our purpose in the game through the avatars, and skillful masculine and skillful feminine action is encoded in a way that reinforces gender bias.

Background Relations

We can address a final category in human-technics relations from Ihde (1990)—background relations—in which the technology forms an environment to other interactions but disappears entirely from the user experience. Air conditioning, controlled by a thermostat, is a clear and paradigmatic example. Here, already, we have a good example for our consideration as well! Air conditioning is designed to disappear from our experience, but is foregrounded much more for women than men. Expected business attire for men—long sleeves, long pants, layers—breathes less and is less drafty than women’s skirts and blouses. This is all the more true when we look at more traditional business attire, as was expected of men working in offices when air conditioning first came to be standard equipment. This, even without considering biological differences (Byrne et al. 2005)! It should not be surprising then that the expectation of what an office feels like would be tailored to men’s embodiment and clothing, as a recent study found (Kingma and Lichtenbelt 2015)—although it may be quite surprising to consider how frequently we fail to recognize that this constitutes an unearned advantage, instead simply thinking that it seems like women are always cold.

To make this clear through phenomenological variation, consider men’s experience of a workplace thermoneutral for the majority of women’s clothing and metabolic rates: many men would be sweaty and unkempt, unbuttoned and unprofessional. Instead, it is women who are subjected to a meteorologically hostile work environment, and we have for the most part accepted that it’s women’s responsibility, individually, to bring sweaters and space heaters rather than our collective responsibility to provide an office environment hospitable—or failing that, equally uncomfortable—for all employees.

Innumerable sorts of online algorithms also present background relations that have discriminatory effects. These algorithms determine what we see online in a way that is entirely inaccessible to user knowledge or control—and, further, most users are not even aware that what they see is different in an important way from what others see.

Eli Pariser (2012) has written about the “filter bubble” as a way of understanding the customized internet; Cass Sunstein (2009) has addressed similar effects through his idea of the “Daily We,” in which we seek out information that fits with our existing views, leading to “ideological echo chambers” (Pew Research Center 2014) and “information cocoons” (Sunstein 2009: 44), resulting in an online environment that Pariser calls “a city of ghettos” (2012)—all largely without the user’s awareness. For the most part, and in spite of the racial analogy of the ‘ghetto,’ this discussion has mostly focused on the way that political opinions have become retrenched and insularized rather than the way that it creates insularity within groups subject to discrimination and creates blindness to those groups in ‘average,’ normatively privileged users.

Trackers on webpages follow our IP addresses and create profiles that are used by advertising firms in real-time bidding to determine which ads are served to us. To see the action of these trackers, go to Amazon and search for something you’re not interested in—sump pumps, homeopathic remedies, or blouses, perhaps—and watch how related ads follow you around

the internet in the next few days. These ad buys are informed not only by browsing history, but also by other data obtainable through your IP address: for example, your zip code, and the median house price and voting record for your zip code. The background to what we search for and read becomes tailored to not only the online environment we create but to the offline environment as well. These have racial and gendered aspects, of course, creating an invisible online ghettoization of women and minorities and an invisibly White and male internet for White male users.

Google search results are similarly customized using IP address and browsing history—with search history as well, if the user is signed in and allows Google to do so. The purpose of this customization of results is to deliver the information that users will find most relevant to their interests, but this clearly threatens to create racialized and gendered distortions, strongly compounded by the trust that users place in Google to provide relevant and objective information.

The extent and impact of the customized internet is difficult to study due to the complexity of these information ecosystems and their unavailability to public view, but it is clear that the impact can be significant. One study found that users identified algorithmically as female received “fewer instances of an ad related to high paying jobs” (Datta et al. 2015); another found “statistically significant discrimination in ad delivery based on searches of 2184 racially associated personal names,” where searches for statistically Black names such as DeShawn rather than statistically White names such as Geoffrey resulted in a greater proportion of ads suggesting criminal arrest records (Sweeney 2013). An invisible return of redlining also seems to be implied by this process, where those living in minority-majority zip codes may have an online environment that emphasizes e.g., payday loan services over mortgage services and educational opportunities. A patent acquired by Facebook even allows for assessment of a loan applicant through the Facebook Graph API:

When an individual applies for a loan, the lender examines the credit ratings of members of the individual’s social network who are connected to the individual through authorized nodes. If the average credit rating of these members is at least a minimum credit score, the lender continues to process the loan application. Otherwise, the loan application is rejected. (Lunt 2014)

It is not hard to imagine how this and other applications of social networking site APIs could have discriminatory results, given that online associations tend to emphasize in-group connections by race, sexual orientation, transgender status, and so on. Further, as the Internet of Things continues to expand, identity markers of increasing variety and specificity will be available to these algorithms.

What Should Be Done?

We are all subject to various microaggressions at the hands of our tools—nearly every artifact designed to interface with persons must make significant assumptions about that person’s attributes in order to withdraw into their experience. These rise to the level of discrimination only when they are persistent and significant in effect, or when they play into and reinforce existing discriminatory structures or attitudes. While One is not 6'4", microaggressions to someone of that height amount to little aside from bumping into the occasional beam or chandelier, and having to crouch down or lean over to wash one’s hair in the shower. Left-handedness makes many things more difficult and uncomfortable, and accommodating artifacts, like left-handed scissors and guitars, are manufactured in order to address these difficulties. The persistence of

these microaggressions and the seriousness of the difficulties they present is, however, incomparably greater in the case of someone who moves around on wheels rather than legs, and these microaggressions would produce a discriminatory effect even in the absence of ableist prejudice.

In most cases, we can't design to be responsive to every person, body, identity, and circumstance, and most of the time designing mostly for most people is good enough. It is clear that adhesive bandages available in only a single version of a 'flesh' color is insufficient, and that ten different shades is probably more than necessary to avoid the negative discriminatory effect—but it's difficult to say how or even whether this should be applied to kitchen countertop design. What counts as a reasonable accommodation also likely depends strongly on individual applications, what's at stake, and what cost and effort accommodations will demand. A strong effort has been made to accommodate people in wheelchairs in architecture, for example, and diaper changing tables—a cheap solution to a smaller problem—are increasingly common in men's bathrooms. Making virtual reality systems equally accessible to women is not currently recognized as so incumbent upon developers and companies, and insofar as such systems are used for entertainment only, it may well be sufficient for companies to say "we design for our target market," and to treat the exclusion of many women as a marketing choice rather than an issue of equity. When we look at the use of virtual reality in military treatment of posttraumatic stress disorder, however, it is more clear that providing unequal medical treatment to female veterans is a serious inequity, resulting in significant discrimination. We can imagine, similarly, that as customized pharmaceuticals and personalized medicine continue, the current research and development funding bias towards the diseases of the wealthy will result in increasing health care disadvantages for the poor—and, if such new forms of medicine attach to genetic differences across race or gender, other forms of health care discrimination may result as well. A further emerging set of issues is presented by a move to decentralized service provision under the 'sharing economy': services such as Uber and AirBnB do not well support persons with disabilities, with some reports of service animals being placed in trunks and blind persons being injured by drivers refusing them service (Heideman 2014), and business models that put provision of architectures of accommodation upon independent contractors are not well organized to ensure that all customers have equitable access.

In background relations, however, the problem isn't so much a lack of responsivity as it is an excess of 'accommodation' in the form of targeting and customization. In either kind of case, though, the problem results from the Latourian delegation where the technical system enforces a cultural norm in an inflexible way through a causal system.

So, no general policy can be recommended except that designers and programmers be aware of how these disparities arise, and design in a way which minimizes them as appropriate to the particular tool and task at hand. What is needed is a process of diversity impact assessment, akin to environmental impact assessments already regularly conducted in some areas of technological design and implementation. Having greater diversity in technology and engineering fields would help, of course—and would be desirable for its own sake—but this is not something that should be beyond the abilities of straight White able-bodied Christian cis men! A good number of these issues can be addressed by stepping back periodically and asking how design choices will differentially impact people and communities different from one's own.

A good number of other problems, however—such as personalized information spaces, social organization through calendars, and established gender roles—are so diffuse and multiply grounded in social norms and technical systems that they may be nearly insoluble. It is an existential rather than a technical problem that the One that forms the averageness with relation to which we live is something from which *all* of us deviate and are excluded from in one way or another.

Note

1 I use this example because it is the most famous of Winner's cases from this influential example, and it does illustrate his point effectively. Several authors, however, have presented strong arguments that the case is historically and factually erroneous, e.g., Joerges (1999a), Woolgar and Cooper (1999), and Joerges (1999b).

References

- boyd, d. (2014). Is the Oculus Rift Sexist? *Quartz*, March 28. Available at: <http://qz.com/192874/is-the-oculus-rift-designed-to-be-sexist/> [Accessed August 10, 2015].
- Byrne, N. M., Hills, A. P., Hunter, G. R., Weinsier, R. L., and Schutz, Y. (2005). Metabolic Equivalent: One Size Does Not Fit All. *Journal of Applied Physiology*, 99, 112–119.
- Datta, A., Tschantz, M. C., and Datta, A. (2015). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *Proceedings on Privacy Enhancing Technologies*, 2015(1), 92–112.
- Du Bois, W. E. B. (1994). *The Souls of Black Folk*. New York: Dover Publications. (Original work published 1903)
- Dyer, R. (1997). *White*. London: Routledge.
- Floridi, L. (2014). *The Fourth Revolution*. Oxford: Oxford University Press.
- Heidegger, M. (1996). *Being and Time*, trans. J. Stambaugh. Albany: SUNY Press. (Original work published 1927)
- Heideman, E. (2014). Uber and Airbnb Leave Disabled People Behind. *The Daily Beast*, October 4. Available at: www.thedailybeast.com/articles/2014/10/04/uber-and-airbnb-leave-disabled-people-behind.html [Accessed August 10, 2015].
- HP Computers Are Racist (2009). video, wzamen01, December 10. Available at: www.youtube.com/watch?v=t4DT3tQqRM [Accessed August 10, 2015].
- Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth*. Bloomington, IN: Indiana University Press.
- Joerges, B. (1999a). Do Politics Have Artefacts? *Social Studies of Science*, 29(3), 411–431.
- Joerges, B. (1999b). Scams Cannot Be Busted: Response to Steve Woolgar and Geoff Cooper, 'Do Artefacts Have Ambivalence? Moses' Bridges, Winner's Bridges and Other Urban Legends in STS. *Social Studies of Science*, 29(3), 450–457.
- Kingma, B., and Lichtenbelt, W. v. M. (2015). Energy Consumption in Buildings and Female Thermal Demand. *Nature Climate Change*, 5(9), 1054–1056.
- Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. In W. Bijker and J. Law (eds.) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge: Cambridge University Press.
- Latour, B. (1999). A Collective of Humans and Nonhumans: Following Daedalus's Labyrinth. In B. Latour (ed.) *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
- Lunt, C. (2014). *Authorization and Authentication Based on an Individual's Social Network*. U.S. Patent 8782753 B2.
- McIntosh, P. (1989). White Privilege: Unpacking the Invisible Knapsack. *Peace and Freedom Magazine*, July/August.
- Mullen, J. (2015). Google Rushes to Fix Software That Tagged Photo With Racial Slur. *CNN*, July 2. Available at: www.cnn.com/2015/07/02/tech/google-image-recognition-gorillas-tag/. [Accessed August 10, 2015].
- Pariser, E. (2012). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin Books.
- Pew Research Center. (2014). *Political Polarization in the American Public: How Increasing Ideological Uniformity and Partisan Antipathy Affect Politics, Compromise and Everyday Life*, June 12. Available at: www.people-press.org/2014/06/12/political-polarization-in-the-american-public/ [Accessed August 10, 2015].

- Roth, L. (2009). Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity. *Canadian Journal of Communication*, 34(1). Available at: <http://cjc-online.ca/index.php/journal/article/viewArticle/2196>
- Smith, D. (2013). 'Racism' of Early Colour Photography Explored in Art Exhibition. *The Guardian*, January 25. Available at: www.theguardian.com/artanddesign/2013/jan/25/racism-colour-photography-exhibition [Accessed August 10, 2015].
- Sunstein, C. (2009). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
- Sweeney, L. (2013). Discrimination in Online Ad Delivery. *ACM Queue*, 11(3). doi:10.1145/2460276.2460278.
- Wilcox, P. (1970). *White Is*. New York: Grove Press.
- Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109(1), 121–136.
- Woolgar, S., and Cooper, G. (1999). Do Artifacts Have Ambivalence? Moses' Bridges, Winner's Bridges and Other Urban Legends in S&TS. *Social Studies of Science*, 29(3), 433–449.

Video Games and Ethics

Monique Wonderly

In 2012, the global video game market was valued at approximately 79 billion dollars, and according to some estimates, its value is expected to exceed 110 billion dollars by the end of 2015 (Gartner 2013). In the United States alone, more than 155 million people play video games, and more than half of all U.S. households own a dedicated gaming console (ESA 2015). Video games constitute one of the most thriving, pervasive, and popular forms of media in the world.

The mass appeal of video games doubtless has been fueled by impressive advances in computer and console gaming technologies—including accelerated graphics, motion-sensitive control devices, and the sophisticated design of expansive ‘open’ game worlds. Players are increasingly afforded richer, more realistic options for exploring and interacting with their virtual environments. Yet, while these features may make video games more enjoyable, they also make certain questions concerning the *ethics* of playing video games more salient. For example, consider the player who repeatedly thrusts the controller forward in order to simulate the stabbing murder of a computer-controlled human character who appears to scream, bleed, and writhe in pain. Some would deem the player’s actions in this scenario morally wrong, while others would insist that they have done nothing at all ethically objectionable.¹

At first glance, it is unclear whether and how there can be anything morally significant about video game play that does not directly involve other human beings. The player in our hypothetical example isn’t engaging with any other *actual* people. On what grounds, then, might their actions be morally problematic? This question has generated a great deal of debate among theorists who’ve engaged with the topic. In what follows, I present and critically analyze key aspects of this debate. Following the trend in the relevant literature, I focus on the moral status of playing games that are thought to feature excessive or loathsome violence, including sexual violence.

Morally Controversial Video Games

Over the past forty years, a number of video games have engendered controversy because their content was perceived as excessively violent or otherwise offensive. In 1982, Mystique released a game for the Atari 2600 console called *Custer’s Revenge*, the object of which is to have sex with a bound Native American woman. Though the game’s designer denied the accusation, many have denounced the game for depicting rape (Ocala Star-Banner 1982). Two games that explicitly permit virtual rape include Illusion’s PC titles, *Battle Raper: Hyper Rea-laction* (2002) and *RapeLay* (2006). Unsurprisingly, despite being marketed as ‘adult-themed’ games, all three elicited moral outrage from the general public (Peckham 2010).

While relatively few mainstream video games permit sexual interaction between characters and fewer still permit sexual *violence*, violence in other forms has long been a staple of the video game genre. In 1976, the Exidy arcade game, *Death Race*, in which the object is to earn points by using a small vehicle to run down stick figures, drew national attention in the United States, with some critics characterizing the game as “gross,” “sick,” and “morbid” (Young 1976). Twenty years later, controversy arose around two other games featuring vehicular violence: *Carmageddon* (Interplay 1997) and *Grand Theft Auto* (DMA Design/Rockstar North 1997). Both games rewarded players for running over pedestrians and smashing into other vehicles. The *Grand Theft Auto* (GTA) series gained further notoriety when the public got wind of a popular strategy among players of *Grand Theft Auto: Vice City* (Rockstar North 2002). Players could restore their health by paying to have sex with a prostitute whom they could then kill directly after sex in order to take back the funds that they spent on her services.

Grand Theft Auto V (Rockstar North 2013) also permits the virtual murder of prostitutes—and that of many other nonthreatening game characters besides—but what many have deemed more disturbing is that a particular mission *requires* the player’s character to engage in torture. As investigative journalist Simon Parkin explains,

the “24”-esque scene, which requires players to rotate the game controller’s sticks in order to tug out the victim’s teeth with pliers, has inspired debate—not only over its artistic merit but also over whether such distressing interactions have any place in video games.

(Parkin 2013)

The GTA games are generally considered *ultra-violent* video games, a category intended to pick out games that feature graphic depictions of cruelty and repetitive loathsome violence toward human beings (Standler 2007; Media Coalition 2007). A prime exemplar of ultra-violent video games is Rockstar Games’ 2003 release, *Manhunt*.

In *Manhunt*, the main character is forced to work for a snuff film maker who demands the grisly murder of victims using weapons ranging from hammers to plastic bags. The game’s sequel, *Manhunt 2*, appeared on the market in 2007, shortly after the release of Nintendo’s Wii console system. Gamespot staff reviewers explained how the Wii’s motion-sensitive remote impacted game play:

The big hook to the action is the way the controller is used during the various kills. You’ll now have to actively follow onscreen motions that approximate your actions. Throwing a chair? Hold the Wii Remote and analog stick as if you’ve got the chair in your hands and move it accordingly. Stabbing someone? Jab that Wii Remote.

(Cocker and Torres 2007)

Both games have been widely criticized for morally objectionable content, and each has been banned in certain venues (BBC News 2014).

Thus far, we have established that video games with content perceived as excessively violent are prone to public protest and moral criticism. What we have not yet established is whether (and if so, why) such protest and moral criticism might be *justified*. The fact that many people find these games offensive is not enough. Even the most heinous act of violence, if performed by a player character in a game world, cannot directly injure any actual human beings. What, then, is the moral harm in playing video games that enable players to simulate acts of murder, rape, or torture against mere computer-controlled characters?

Let’s consider how ethicists have attempted to address this important question.

Approaches to Assessing the Moral Significance of Playing Violent Games

Behavioral Effects of Violent Video Games

One way to vindicate the claim that violent video games are morally problematic would be to show that playing such games impacts one's moral behavior outside of the gaming environment. Some have suspected that playing violent video games can cause one to violate one's moral duties and/or to behave violently toward others. Let's begin with the former possibility, that video games can cause one to violate one's moral duties. Though presumably, we have no moral duties to virtual characters, we do have certain moral obligations to our fellow human beings and (according to some ethicists) to ourselves. If playing excessively violent video games violates—or makes it significantly more difficult to satisfy—such obligations, then we have a *prima facie* reason for thinking that playing those sorts of games is ethically objectionable.

Eighteenth-century German philosopher Immanuel Kant famously argued that as human beings, we have a moral duty to act so as to respect ourselves and other rational agents (Kant 1997a: 38). As such, we should avoid performing actions that might constitute, or cause, the degradation of our own persons or the disrespect of other autonomous, rational beings. Some theorists have suggested that a Kantian argument can be used to elucidate a moral problem with playing violent video games. David Waddington identifies two “Kantian grounds” on which playing a violent game might be deemed morally problematic. First, when people indulge in a violent video game, they might debase themselves by “acting cruelly.” Second, insofar as video game characters are analogs to human beings, failing to treat them with respect might make us less likely to perform our duties to actual people (Waddington 2007: 125). Let's consider each point in turn.

As Waddington explains, Kant considered certain vices, such as lying and avarice, violations of one's duties to oneself, and he would likely regard behaving “cruelly” in a similar vein (Waddington 2007: 124). Of course, it isn't entirely clear that one can behave cruelly in the context of playing a video game that does not involve other sentient beings. Waddington himself acknowledges that this suggestion is a contentious one, and Marcus Schulzke outright rejects it. According to Schulzke, “there is nothing worthy of being called ‘cruelty’ in video games because the characters are not capable of feeling pain or suffering” (2010: 128). To my mind, Schulzke's conclusion is a bit hasty. It might be possible to behave cruelly even where one knows that one's actions will not cause suffering—e.g., where one revels in causing representations of cruelty for cruelty's sake—but let's set aside this possibility for a moment, and consider another. Even if players do not *behave cruelly* when engaging in video game violence, perhaps they debase themselves in other ways.

For example, one might argue that while indulging in violent video games is not strictly speaking *cruel*, it degrades oneself insofar as it prevents one from engaging in more productive, morally worthwhile activities. This would be especially problematic given that many who play video games (violent or otherwise) tend to invest excessive amounts of time in the activity—often to the point of pathology. Pathological video game play—and indeed, video game *addiction*—is a growing problem in some countries (Chiu et al. 2004; Anderson and Warburton 2012). Notably, there are many cases of players becoming so immersed in video games that they jeopardize their education, their jobs, and even their personal relationships (Whitcott 2011; Brey 2008). There are also cases of players who—refusing to break for self-care—become dangerously dehydrated, starved, and exhausted while playing video games, sometimes leading to death (Hunt and Ng 2015).

Importantly, though, while this extreme brand of self-neglect would surely be morally problematic by Kantian standards, the majority of persons who play violent video games do so while maintaining careers and relationships, and many regard their gaming time as a *form* of self-care rather than a diversion from it. The fact that a small minority of gamers pathologically overindulge in violent video games gives us little reason to think that playing violent video games is inherently self-debasing.

Even if playing violent video games does not violate duties to oneself, one might suggest that playing such games makes it more difficult to satisfy duties to others. The second Kantian argument that Waddington appeals to suggests this possibility. In his *Lectures on Ethics*, Kant explained that while we have no duties to nonhuman animals, we should take care in how we treat them as “he who is cruel to animals becomes hard also in his dealing with men” (Kant 1997b: 212). According to Waddington, if animals can be considered an analog of humanity, then perhaps virtual characters can as well, and so behaving maliciously or otherwise disrespectfully toward virtual characters might incline us toward mistreating actual persons (Waddington 2007: 125).

One criticism of this approach is that it assumes a dubious analogy between our relationship to nonhuman animals and our relationship to virtual human characters. Insofar as Kant’s argument is plausible, it may be because human beings and nonhuman animals are both living, sentient creatures. Behaving cruelly toward sentient nonhuman creatures might plausibly incline one toward mistreating persons. Of course, virtual characters are not sentient creatures, but rather, *representations* of human beings. Schulzke, for example, argues that because virtual characters lack “life and autonomy,” they are akin to photographs and at best, “superficial analogs” to humans (Schulzke 2010: 128).

Schulzke’s argument here may be suspect. First, as some theorists have noted, virtual human characters may be more realistic analogs to actual human beings than are nonhuman animals (Brey 1999: 9). While virtual characters are *in fact* lifeless entities, players interact with them in the game world as though they are alive. When players act upon virtual characters, those characters appear to respond, and in much the same way that we would expect an actual live human being to do. This, it seems, is part of the fun. In this way, it seems reasonable to think that engagement with virtual human beings could constitute a kind of *practice* for engaging with actual human beings—in much the same way that using flight simulation software to fly a virtual airplane can constitute a kind of practice for flying an actual plane. In any case, acting upon virtual characters is certainly not like acting upon photographs. Our orientations toward virtual characters, and our language about them, suggest as much. Consider, for example, that when a player uses his avatar to strike a computer-controlled human character, we might expect him to say something like, “Did you see what happened when I punched that guy?” If he uttered the same question after punching a photograph, we would think it more than a little odd.

The Kantian suggestion that disrespectful treatment of human analogs might cause one to “become hard in his dealings with men” raises difficult questions. What would it mean to treat virtual characters harshly or disrespectfully? And what exactly constitutes ‘hardness’ in our dealings with persons? Fortunately, one needn’t appeal to Kant—or to the notions of duty or respect—in order to show that playing violent video games can impact one’s moral behavior outside the game environment. One might instead simply attempt to show that playing such games causes individuals to behave more violently (or to behave violently more often) in the real world. If a consequence of playing violent video games is an increase in actual violence, then—assuming the benefits of playing such games do not outweigh its costs—there is good reason to think that playing excessively violent video games is morally objectionable.

Can playing violent video games dispose one toward actual violence? Some people certainly seem to think so. There is no shortage of cases in which an outraged public has implicated

violent video games as causal factors in violent crime. After Eric Harris and Dylan Klebold shot more than thirty people (killing thirteen) in what came to be known as the Columbine High School Massacre, many were quick to point out that the pair regularly played *Doom* a first-person shooter game (Brey 2008: 378). In 2004, a teenager who frequently played *Manhunt* stabbed and bludgeoned another child to death. The victim's parents partially blamed the murder on the game, adducing similarities between his actions and the killing techniques used in *Manhunt* (Thorsen 2004). In 2004 and 2006, attorneys filed suits against Rockstar Games, alleging that *Grand Theft Auto: Vice City* contributed to the murderous actions of Devin Moore and Codey Posey respectively (Tuscaloosa News 2006: 4B). Other tragedies for which video games have been blamed include the Virginia Tech Massacre, the Sandy Hook Murders, and the Washington Naval Yard Shootings. Various sources claimed that the perpetrators of these actions all played violent video games (Kain 2013).

Of course, even if the aforementioned young murderers were frequent players of violent video games, this by itself means fairly little. As I argued in an earlier work:

Incidents in which video game players commit violent acts are no doubt tragic, but appealing to such cases as evidence for the legal or moral culpability of violent video games may be problematic. In the first place, millions of children play violent video games, so it is statistically probable that many juvenile offenders will also be players of violent video games strictly as a matter of chance. Second, in most cases which attempt to link violent video games to teenage violence, there are other common threads which appear to be more causally relevant, such as abuse or depression.

(Wonderly 2008: 4).

More to the point, the vast majority of people who play violent video games do not commit violent crimes, so we should resist the impulse to assume that when players of violent video games commit actual violence, that they necessarily do so *because* of the games.

Psychologists and sociologists have become increasingly interested in determining whether there is a credible link between violent video game play and real-world violence. Some empirical studies have suggested a positive relationship between playing violent video games and violent behavior (Anderson and Dill 2000). Other studies, however, appear to contradict these findings (Ferguson 2007). Recently, the American Psychological Association (APA) assembled a task force to conduct a meta-analysis of the studies available on the topic, and the task force concluded that there is insufficient evidence to suggest a causal link between violent video game play and violent criminal behavior (APA 2015: 26). Interestingly, however, the APA did confirm a positive causal relationship between violent video game play and aggression. According to the report:

violent video game use has an effect on aggression. This effect is manifested both as an increase in negative outcomes such as aggressive behavior, cognitions, and affect and as a decrease in positive outcomes such as prosocial behavior, empathy, and sensitivity to aggression.

(APA 2015: 26)

Importantly, not all aggression—or aggressive behavior—translates to actual violence, so the extent to which violent video game play can cause genuinely dangerous behavior remains unclear. What is notable, though, is that even if violent video game play does not directly cause one to behave immorally, the APA analysis provides some evidence that such play can have an undesirable impact on morally significant *attitudes*.

If playing violent video games does in fact negatively impact one's moral (or morally relevant) attitudes, then this also constitutes a reason to think that such play is ethically suspect. In the following section, I review what philosophers, psychologists, and other theorists have had to say on the issue.

Psychological Effects of Violent Video Games

Some theorists have suggested that playing violent video games can be problematic—not necessarily because they directly cause immoral *behaviors*—but rather, because they impact certain of our attitudes in morally pernicious ways. For example, theorists have expressed worries that playing violent video games might cause increases in aggressive cognitions and affect, and decreases in sensitivity to actual violence and empathy.

As noted earlier, a recent APA analysis concluded that among the effects of violent video game use are increases in aggressive cognition and aggressive affect. Examples of aggressive cognition included thoughts about the world being a hostile place, dehumanization, and proviolence attitudes (APA 2015: 10). Aggressive affect was marked by increased feelings of hostility—often accompanied by insensitivity to the distress of others (APA 2015: 10–11).

The worry that violent video game exposure can ‘desensitize’ one to actual violence (or moral atrocity more broadly) has become a prevalent theme in the psychological and philosophical literature on violent video games. Carnagey, Anderson, and Bushman (2007) conducted one of the first studies to experimentally examine the link between violent video game play and physiological desensitization. According to their research, participants who were randomly assigned to play a violent video game, even for just twenty minutes, had relatively lower heart rates and galvanic skin responses while watching footage of real violence than did those randomly assigned to play a nonviolent video game. The research team concluded, “The present experiment demonstrates that violent video game exposure can cause desensitization to real-life violence. . . . It appears that individuals who play violent video games habituate or ‘get used to’ all the violence and eventually become physiologically numb to it” (Carnagey et al. 2007: 495). Subsequent studies measuring violent video game players’ somatic and neural responses to violent stimuli have yielded similar results (see, for example, Arriaga, Monteiro, and Esteves 2011; Engelhardt et al. 2011).

Interestingly, reports of military personnel also provide some support for the idea that violent video games can be used to desensitize individuals to actual violence. Lieutenant Colonel David Grossman notes that the U.S. military employs violent video games to train soldiers. According to Grossman, there is a natural aversion to killing one’s own kind, so troops must be desensitized and conditioned in order to become willing and proficient killers (Grossman 1998). There are also first-person reports from soldiers in the field who’ve indicated that violent video game play helped make it psychologically easier to fire on enemies. Sergeant Sinque Swales recounts one of the first times that he shot an enemy: “It felt like I was in a big video game. It didn’t even faze me, shooting back. It was just natural instinct. *Boom! Boom! Boom! Boom!*” (Vargas 2006). Swales reported being an avid fan of the first-person shooter game *Halo 2* and the military-themed *Full Spectrum Warrior*—a game developed with help from the U.S. Army.

A number of theorists have expressed ethical concerns about the relationship between emotional desensitization and playing violent video games. Thomas Nys, for example, suggests that the “moral fishiness” of playing (some kinds of) violent video games might lie “in their willing desensitization against practices such as rape, murder, or general mischief” (Nys 2010: 85). Waddington expresses a related worry, suggesting that as violent video games increase in verisimilitude, it may become difficult to distinguish between real and simulated transgressions,

and as a result, we would come to “devalue the idea of wrongness” (Waddington 2007: 127). Others have focused on the relationship between violent video game play and reduced empathy.

In an earlier work, I adduced empirical research in support of the view that playing (some forms of) violent video games might damage one’s empathic faculties (Mathiak and Weber 2006; Funk et al. 2004; Bartholow et al. 2005). I argued that if playing such games *does* negatively impact one’s capacity for empathy, then doing so might not only decrease our emotional reactivity to actual violence—nor merely cause us to devalue the idea of wrongness—but it might contribute to the dissolution of our abilities to make moral judgments in general (Wonderly 2008). This is because, on many accounts, our capacities for empathy play an important role in our abilities to glean moral knowledge and to make moral assessments (e.g., Hume 2005; Smith 2009; Hoffman 2000; Slote 2004).

Some theorists have challenged the view that violent video game play can impact one’s morally relevant attitudes in the ways suggested. Schulzke, for example, posits that the arguments put forth by Waddington and Wonderly rely on a dubious analogy between harming virtual characters and harming actual human beings. According to Schulzke, it is unreasonable to suppose that the mere physical resemblance of virtual characters to actual humans is sufficient to damage our abilities to distinguish between real transgressions and simulated ones, or again, to empathize with our fellow human beings (Schulzke 2010: 134). Importantly, though, neither Waddington nor Wonderly argue for this view. While the advent of virtual environments that look (and feel) more realistic would likely exacerbate the potential effects that Waddington and Wonderly point to, one needn’t think that the cause of either problem would be *reducible* to mere physical resemblance. Other contributing factors might include, for example, the richness and complexity of interactions available to the player, the frequency with which specific types of virtual actions are repeated, and the in-game consequences of performing the virtual acts in question.

Schulzke is also skeptical about the empirical research that suggests a relationship between violent video game play and reduced empathy and emotional sensitivity. Schulzke argues that studies purporting to show such a relationship are often plagued by methodological flaws and research biases. He also argues that the results of such studies are both prone to misinterpretation and must contend with other research that suggests an opposite conclusion (2010: 133). Philip Brey and Garry Young have expressed similar worries (Brey 2008: 378; Young 2013: 33).

The comprehensive metareview recently conducted by the APA task force might help to assuage some of these concerns. According to its report, the available research demonstrates “a consistent relation” between violent video game use and “heightened aggressive affect” along with “reduced empathy and sensitivity to aggression” (APA 2015: 18, 26). Of course, one must take care in interpreting the results too strongly. They do not conclusively prove that violent video game play elevates aggressive attitudes or reduces empathy or sensitivity to violence. The results are, however, quite suggestive. They indicate that the *best available* empirical evidence on the topic suggests that violent video game use can, and often does, impact morally significant aspects of our psychology in potentially pernicious ways.

If playing violent video games can have this kind of impact on one’s psychology, then this would be significant for at least two reasons. First, the psychological effects noted earlier could cause real, if subtle, negative changes in behavior. Recall that while the APA task force concluded that there is insufficient evidence to establish a link between playing violent video games and criminal violence, it is satisfied that playing such games can increase some forms of aggressive behavior (APA 2015: 26). Second, even where such psychological changes do not directly translate into immoral *actions*, increased aggressive affect and cognition and reduced empathy and sensitivity to violence may nonetheless negatively impact one’s *character*. This

is important from an ethical standpoint, as morality is concerned not only with what kind of actions we perform, but also with what kind of people we *are*.

Violent Video Games and Character

The normative approach to ethics broadly known as *virtue ethics*, puts the notion of character—rather than action—at the forefront of moral theory. Virtue ethicists tend to focus on the question, “What kind of person should I be?” On some views, one can use elements of virtue ethics to evaluate actions. For example, actions that enhance one’s virtue or character might be considered good on that account, whereas actions that inculcate vice and thereby harm one’s character might be deemed bad or wrong. Many have advocated adopting a virtue-theoretical framework in order to morally evaluate violent video games (McCormick 2001; Coeckelbergh 2007, 2011; Sicart 2009a; Sicart 2009b).

As Aristotle’s theory of virtue is the most popular exemplar of virtue ethics, theorists often draw on an Aristotelian account of character development in order to explain why playing violent video games might be morally problematic. Aristotle posited that the development of virtues—which on his conception, consist in intellectual, emotional, and social *skills*—is central to human flourishing or *eudaimonia*. Ethical virtues, or virtues of character such as temperance, are cultivated through practice (Aristotle 2000; Kraut 2014). Through proper upbringing and repetition of virtuous habits, we develop the sorts of characters that are conducive to human well-being.

Some theorists have argued that playing violent video games is inimical to virtue because in playing such games, players practice and reinforce morally vicious habits. Matthew McCormick, for example, suggests that what we do wrong when we “pull the virtual trigger” is “reinforce virtueless habits and make it harder for the individual to reach eudaimonic fulfillment” (McCormick 2001: 286). Mark Coeckelbergh endorses a similar view. On his account, by “training moral insensitivity,” and inhibiting the development of empathy, playing violent video games could prevent us from becoming “virtuous, flourishing human beings” (Coeckelbergh 2007: 230; 2011: 94–95). The upshot of this approach is that even if playing violent video games does not directly cause immoral behavior, it nevertheless may damage one’s moral character.

Predictably, not all theorists agree with this view. According to some, playing certain types of violent video games might actually be virtuous. Aristotle’s theory is not the only model of virtue ethics, and it is possible that other virtue theorists would consider some instances of violent video game play praiseworthy. Some, for example, consider Friedrich Nietzsche a virtue ethicist (Brobjer 2003; Daigle 2006). Nietzsche endorsed virtues such as courage, creativity, and strength of will. Playing violent video games might well enhance these virtues (Wonderly 2008: 4–5; Young 2013: 94–95). Marcus Schulzke argues that Aristotle himself suggested that not all violence is irreconcilable with virtue—some virtues being “exemplified in combat,” so even an Aristotelian virtue ethicist needn’t object to violent video games as such (Schulzke 2010: 131).

This line of argument brings to the fore an important point: violent video games vary widely in terms of the particular kinds of virtual acts that they permit or require from players. Taking down an enemy combatant in a first-person shooter game, and stalking and torturing an innocent victim in a stealth-based survivor horror game, if they impact one’s character at all, may do so in very different ways (and to very different degrees). What is needed is a more informative account of the conditions under which playing violent video games impede or diminish virtue.

Interestingly, some philosophers adopt a virtue-theoretic approach not to argue that playing video games can harm one’s character, but that certain reactions to particular kinds of games

might *reflect* poor character in the player. Stephanie Patridge, for example, suggests that some video games feature content that, when viewed against the historical and social backdrop of one's society, has "incorrigible social meaning" and that our responses to such meanings "bear on evaluations of our character" (Patridge 2011: 304). Patridge here has in mind video games like *Custer's Revenge*. She explains that *Custer's Revenge* invites the player to be entertained by a representation of a rape of a Native American woman, but intuitively, one should not enjoy representations like that (2011: 306). In a properly informed and appropriately sensitive individual, such representations would call to mind actual atrocities against unjustly targeted groups—in this case, women and minorities—that should preclude enjoyment.

In a later work, Patridge employs this notion to address what has been dubbed the "Gamer's Dilemma." In brief, the Gamer's Dilemma asks, why, if (as many have claimed) virtual murder is morally acceptable because it harms no one in the actual world, should we not regard virtual pedophilia as equally benign for the same reason (Luck 2009)?² While Patridge does not purport to decisively solve the dilemma, she thinks that elements of her view can explain why virtual child pedophilia might elicit reactions that some other violent video games do not. She explains that a game that invites one to sexually assault a character simply because it represents a child might call to mind actual child sexual assault victims who were similarly targeted because of their youth. This association might (and likely should) make it more difficult—if not impossible—to enjoy the game (Patridge 2013: 32). We would expect some, but not all, nonsexual, violent video games to elicit similar reactions—in particular, games that are thickly laden with social meaning, e.g., a "lynching game" (Patridge 2013: 33).

There are also others who've suggested that one can have morally inappropriate responses to certain types of virtual content, such as enjoying virtual child pornography or becoming sexually aroused by virtual depictions of rape (Veber 2004; Gooskens 2010). In these cases, one might think it reasonable to suspect that the subject of these responses has a morally flawed character. What is less clear, however, is whether such reasoning might be extended to players of violent video games more broadly. Do players who become excited or joyful upon employing their character to kill or torture computer-controlled characters reflect poor character on that account? Or might such reactions represent natural, more or less, benign responses to exploring new, socially 'taboo' activities in a harmless virtual environment? These questions are difficult ones that have inspired a great deal of controversy in the literature on ethics and violent video games (see, for example, Young and Whitty 2012; Young 2013).

Addressing the Concerns

Having rehearsed the most prominent approaches to assessing the ethics of playing violent video games, we are now in a position to take stock. Historically, video games featuring content perceived as excessively violent have drawn moral criticism from an indignant (and sometimes, morally outraged) public. Defenders of violent video games have insisted that such criticisms are unwarranted, as committing acts of virtual violence against computer-controlled characters—no matter how heinous or cruel those actions would be if performed in real life—harm no actual people. Theorists attempting to articulate the moral significance of such games have suggested that playing them can (1) inspire immoral behaviors outside the gaming environment, (2) impact certain aspects of our psychology in morally pernicious ways, and/or (3) damage or reflect poor character. These claims, however, have not gone uncontested. Some have asserted that the differences between virtual and actual violence are salient enough to prevent gamers from letting their experiences in video games bleed over into their actual behaviors and attitudes, and they have challenged the empirical research that suggests otherwise. So the question remains, "Where do we go from here?"

On my view, first we must acknowledge that there is real cause for concern. While it is unlikely that playing an hour of *Grand Theft Auto* will turn a person into a murderer, it is perhaps equally unlikely that regularly and repeatedly simulating acts of wanton, graphic violence against virtual human characters will have no negative psychological impact on a sizable percentage of video game players who do so—especially given that many such players are children. Both common sense and a preponderance of the best empirical research available suggest as much.

Second, we should call for more empirical studies in order to better determine the extent of the impact playing violent video games tends to have, the particular features of violent games that tend to cause the most harm, and the specific groups that are most vulnerable to those harms. These studies should be conducted by unbiased experts and subject to review by independent organizations with the means to fairly and accurately assess the quality of the research. Once the relevant information has been obtained, there should be vigorous attempts to disseminate it to video game players, parents of children who play video games, video game developers, legislators, and other interested parties.

Third, when deciding whether to alter or restrict video game content, policy makers should take into account not only the potential harms of playing violent video games, but the potential *positive* effects of playing such games as well. Some have argued, for example, that playing some kinds of violent video games can be cathartic, allowing players to vent aggression and hostility that might otherwise be inflicted on actual people (Brey 1999: 8; Brey 2008: 369; Nys 2010: 86; Schulzke 2010: 133). Also, some violent video games incorporate morally interesting narratives, artful imagery and musical scores, and sophisticated haptical technologies that allow for gaming experiences that are not only pleasurable, but also aesthetically—and perhaps even morally—enriching. Miguel Sicart has argued that mature players can benefit from violent video games that confront them with moral dilemmas, exploiting a tension between their objectives in the gaming world and their preexisting ethical commitments (Sicart 2009a: 199–200; Sicart 2009b: 113–116). Thus, the moral significance of violent video games may extend not only to their potential harms, but also to their benefits as well.

Ethicists have identified several plausible grounds for suspecting that playing violent video games can be morally problematic, but looming questions threaten to undermine the persuasiveness of their arguments. I have argued that as it stands, there is sufficient reason to take modest steps in order to address the concerns that these theorists have raised. These steps include new empirical studies, vigorous attempts to educate the public about the results of such studies, and a decision-making procedure that takes into account both the negative and positive aspects of playing violent video games.

Notes

- 1 One way to support the latter position is to maintain that a player's actions within a video game are *never* appropriate objects of moral assessment. On this view, what gaming characters ought, or ought not, to do is constrained solely by the rules and objectives of the game itself. Of course, this position seems baldly untenable when applied to some genres of video games. Some games represent virtual *communities* in which actual human beings can interact with one another via their personalized avatars. In games such as *World of Warcraft*, *EVE Online*, and *The Elder Scrolls*, players can not only challenge one another in battle, but they can also establish friendships, forge alliances, engage in complex economic transactions, and in some cases, even (virtually) marry other players. In these environments, while individuals cannot physically injure one another, gamers can—and sometimes do—maliciously deceive, harass, and steal from their fellow players, even where such actions are not prescribed by the game. In these gaming contexts, then, it is fairly clear that virtual actions can have real-world ethical implications. For more on the ethical status of actions performed within virtual communities, see Dunn (2012).

- 2 Luck's puzzle has prompted a number of responses. Christopher Bartel, drawing on Neil Levy's work on virtual child pornography, argued that while virtual pedophilia may not harm actual children, it may yet—in virtue of sexualizing inequality—harm actual women (Bartel 2012). A decade earlier, Neil Levy argued that obtaining “equal status” for all women requires a “new sexuality” in which sexual relations are conducted between equals, and by “eroticizing inequality,” virtual child pornography might hinder progress toward this important ideal (Levy 2002: 322). For commentary on Bartel's application of this suggestion to the Gamer's Dilemma, see Luck and Ellerby 2013 and Patridge 2013.

References

- American Psychological Association (APA). (2015). American Psychological Association Task Force on Violent Media Technical Report on the Review of the Violent Video Game Literature. Available at: www.apa.org/news/press/releases/2015/08/violent-video-games.aspx [Accessed August 17, 2015]
- Anderson, C. A., and Dill, K. E. (2000). Video Games and Aggressive Thoughts, Feelings, and Behavior in the Laboratory and in Life. *Journal of Personality and Social Psychology*, 78(4), 772–790.
- Anderson, C., and Warburton, W. (2012). The impact of Violent Video Games: An Overview. In W. Warburton and D. Braunstein (eds.) *Growing Up Fast and Furious: Reviewing the Impacts of Violent and Sexualised Media on Children*. Annandale, NSW, Australia: The Federation Press.
- Aristotle. (2000). *Nicomachean Ethics*, trans. R. Crisp. Cambridge: Cambridge University Press.
- Arriaga, P., Monteiro, M.B. and Esteves, F. (2011). Effects of Playing Violent Computer Games on Emotional Desensitization and Aggressive Behavior. *Journal of Applied Social Psychology*, 41(8), 1900–1925.
- Bartel, C. (2012). Resolving the Gamer's Dilemma. *Ethics and Information Technology*, 14(1), 11–16.
- Bartholow, B., Sestir, M., and Davis, E. (2005). Correlates and Consequences of Exposure to Video Game Violence: Hostile Personality, Empathy, and Aggressive Behavior. *Personality and Social Psychology Bulletin*, 31(1), 1573–1586.
- BBC News. (2014). ‘Sexually Violent’ GTA 5 Banned From Australian Stores.” December 4. Available at: www.bbc.com/news/technology-30328314 [Accessed August 6, 2015]
- Brey, P. (1999). The Ethics of Representation and Action in Virtual Reality. *Ethics and Information Technology*, 1(1), 5–14.
- Brey, P. (2008). Virtual Reality and Computer Simulation. In K. E. Himma and H. T. Tavani (eds.) *The Handbook of Information and Computer Ethics*. Hoboken, NJ: Wiley-Interscience, 361–384.
- Brobjer, T. H. (2003). Nietzsche's Affirmative Morality: An Ethics of Virtue. *The Journal of Nietzsche Studies*, 26(1), 64–78.
- Carnagey, N. L., Anderson, C. A., and Bushman, B. J. (2007). The Effect of Video Game Violence on Physiological Desensitization to Real-Life Violence. *Journal of Experimental Social Psychology*, 43(3), 489–496.
- Chiu, S.-I., Lee, J.-Z., and Huang, D.-H. (2004). Video Game Addiction in Children and Teenagers in Taiwan. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 7(5), 571–581.
- Cocker, G. & Torres, R. (2007). Manhunt 2. *Game Spot*. Available at: www.gamespot.com/wii/action/manhunt2/news.html [Accessed August 10, 2007]
- Coeckelbergh, M. (2007). Violent Computer Games, Empathy, and Cosmopolitanism. *Ethics and Information Technology*, 9(3), 219–231.
- Coeckelbergh, M. (2011). Virtue, Empathy, and Vulnerability: Evaluating Violence in Digital Games. In K. Poels and S. Malliet (eds.) *Vice City Virtue: Moral Issues in Digital Game Play*. Leuven/Den Haag: Acco Academic, 89–105.
- Daigle, C. (2006). Nietzsche: Virtue Ethics . . . Virtue Politics? *The Journal of Nietzsche Studies*, 32(1), 1–21.
- Dunn, J. (2012). Virtual Worlds and Moral Evaluation. *Ethics and Information Technology*, 14(4), 255–265.
- Engelhardt, C. R., Bartholow, B. D., Kerr, G. T., and Bushman, B. J. (2011). This Is Your Brain on Violent Video Games: Neural Desensitization to Violence Predicts Increased Aggression Following Violent Video Game Exposure. *Journal of Experimental Social Psychology*, 47(5), 1033–1036.

- Entertainment Software Association (ESA). (2015). *Essential Facts about the Computer and Video Game Industry*. Available at: www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf [Accessed August 3, 2015]
- Exidy. (1976). *Death Race*.
- Ferguson, C. J. (2007). The Good, the Bad and the Ugly: A Meta-Analytic Review of Positive and Negative Effects of Violent Video Games. *Psychiatric Quarterly*, 78(4), 309–316.
- Funk, J., Baldacci, H., Pasold, T., and Baumgardner, J. (2004). Violent Exposure in Real-Life, Video Games, Television, Movies, and the Internet: Is There Desensitization? *Journal of Adolescence*, 27, 23–35.
- Gartner (2013). Gartner Says Worldwide Video Game Market to Total \$93 Billion in 2013. Available at: www.gartner.com/newsroom/id/2614915 [Accessed August 3, 2015]
- Goossens, G. (2010). The Ethical Status of Virtual Actions. *Ethical Perspectives*, 17(1), 59–78.
- Grossman, D. (1998). Trained to Kill. *Killology Research Group: The Psychological Cost of Learning to Kill*: Available at: www.killology.com/art_trained_killing.htm [Accessed August 10, 2007]
- Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for Caring and Justice*. New York: Cambridge University Press.
- Hume, D. (2005). *A Treatise of Human Nature*. New York: Barnes & Noble, Inc.
- Hunt, K., and Ng, N. (2015). Man Dies After 3-Day Internet Gaming Binge. *CNN.com*, January 19. Available at: www.cnn.com/2015/01/19/world/taiwan-gamer-death/index.html [Accessed August 7, 2015]
- Illusion. (2002). *Battle Raper: Hyper Realaction*.
- Illusion. (2006). *RapeLay*.
- Interplay. (1997). *Carmageddon*.
- Kain, E. (2013). Do Games Like ‘Grand Theft Auto V’ Cause Real World Violence? *Forbes*, September 18. Available at: www.forbes.com/sites/erikkain/2013/09/18/do-games-like-grand-theft-auto-v-cause-real-world-violence/ [Accessed August 7, 2015]
- Kant, I. (1997a). *Groundwork of the Metaphysics of Morals*, trans. and ed. M. Gregor. Cambridge: Cambridge University press.
- Kant, I. (1997b). *Lectures on Ethics*, trans. and ed. P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.
- Kraut, R. (2014). Aristotle’s Ethics. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Summer 2014 ed.). Available at: <http://plato.stanford.edu/archives/sum2014/entries/aristotle-ethics/>.
- Levy, N. (2002). Virtual Child Pornography: The Eroticization of Inequality. *Ethics and Information Technology*, 4(4), 319–323.
- Luck, M. (2009). The Gamer’s Dilemma: An Analysis of the Arguments for the Moral Distinction Between Virtual Murder and Virtual Paedophilia. *Ethics and Information Technology*, 11(1), 31–36.
- Luck, M., and Ellerby, N. (2013). Has Bartel Resolved the Gamer’s Dilemma? *Ethics and Information Technology*, 15(3), 229–233.
- Mathiak, K., & Weber, R. (2006). Toward Brain Correlates of Natural Behavior: fMRI During Violent Video Games. *Human Brain Mapping*, 27(12), 948–956.
- McCormick, M. (2001). Is It Wrong to Play Violent Video Games? *Ethics and Information Technology*, 3(4), 277–287.
- Media Coalition, Inc. (2007). *Legislative Updates*. Media Coalition, Inc., July 26 Available at: www.mediacoalition.org/legislative/ [Accessed August 10, 2007]
- Mystique. (1982). *Custer’s Revenge*.
- Nys, T. (2010). Virtual Ethics. *Ethical Perspectives*, 17(1), 79–93.
- Ocala Star-Banner. (1982). Atari Trying to Halt X-Rated Video Games. Ocala Star-Banner, October 17. Available at: <https://news.google.com/newspapers?nid=1356&dat=19821017&id=2OsTAAAAIBAJ&sjid=-wUEAAAAIBAJ&pg=5332,427806&hl=en> [Accessed July 3, 2015]
- Parkin, S. (2013). “How Evil Should a Video Game Allow You to Be?” *The New Yorker*. Available at: www.newyorker.com/tech/elements/how-evil-should-a-video-game-allow-you-to-be [Accessed August 6, 2015]
- Patridge, S. (2011). The Incurable Social Meaning of Video Game Imagery. *Ethics and Information Technology*, 13(4), 303–312.
- Patridge, S. (2013). Pornography, Ethics, and Video Games. *Ethics and Information Technology*, 15(1), 25–34.

- Peckham, M. (2010). The 15 Most Offensive Games Ever Made. *The Washington Post*, October 4. Available at: www.washingtonpost.com/wp-dyn/content/article/2010/09/29/AR2010092904875.html [Accessed August 2, 2015]
- Rockstar Games. (2003). *Manhunt*.
- Rockstar Games. (2007). *Manhunt 2*.
- Rockstar North. (1997). *Grand Theft Auto*.
- Rockstar North. (2002). *Grand Theft Auto: Vice City*.
- Rockstar North. (2013). *Grand Theft Auto V*.
- Schulzke, M. (2010). Defending the Morality of Violent Video Games. *Ethics and Information Technology*, 12(2), 127–138.
- Sicart, M. (2005). Game, Player, Ethics: A Virtue Ethics Approach to Computer Games. *International Review of Information Ethics*, 4(12), 13–18.
- Sicart, M. (2009a). The Banality of Simulated Evil: Designing Ethical Gameplay. *Ethics and Information Technology*, 11(3), 191–202.
- Sicart, M. (2009b). *The Ethics of Computer Games*. Cambridge, MA: MIT Press.
- Slote, M. (2004). Moral Sentimentalism. *Ethical Theory and Moral Practice*, 7, 3–13.
- Smith, A. (2009). *The Theory of Moral Sentiments*. Indianapolis, IN: Liberty Fund.
- Standler, R. (2007). *State Statutes Regulating Video Games in the USA*. Available at: Dr. R. Standler's Professional Homepage: www.rbs2.com/ssrvrg.pdf [Accessed August 10, 2007]
- Thorsen, T. (2004). Manhunt Blamed in UK Murder. *GameSpot*, July 29. Available at: www.gamespot.com/ps2/action/manhunt/news.html [Accessed August 10, 2007]
- The Tuscaloosa News*. (2006). New Mexico Video Game Lawsuit Like Fayette Case. *The Tuscaloosa News*, September 26. Available at: <https://news.google.com/newspapers?nid=1356&dat=19821017&id=2OsTAAAAIIBAJ&sjid=-wUEAAAAIIBAJ&pg=5332,427806&hl=en> [Accessed July 3, 2015]
- Vargas, J. A. (2006). Virtual Reality Prepares Soldiers for Real War. *The Washington Post*, February 14. Available at: <https://news.google.com/newspapers?nid=1817&dat=20060926&id=IcsyAAAAIIBAJ&sjid=0acEAAAAIIBAJ&pg=6308,4432364&hl=en> [Accessed July 9, 2015]
- Veber, M. (2004). Virtual Child Pornography. *Public Affairs Quarterly*, 18(1), 75–90.
- Waddington, D. I. (2007). Locating the Wrongness in Ultra-Violent Video Games. *Ethics and Information Technology*, 9(2), 121–128.
- Whatcott, A. (2011). Video-Game Addiction Blamed for 15 Percent of Divorces. *Deseret News*, May 31. Available at: www.deseretnews.com/article/700140348/Video-game-addiction-blamed-for-15-percent-of-divorces.html?pg=all [Accessed July 7, 2015]
- Wonderly, M. (2008). A Humean Approach to Assessing the Moral Significance of Ultra-Violent Video Games. *Ethics and Information Technology*, 10(1), 1–10.
- Young, G. (2013). *Ethics in the Virtual World: The Morality and Psychology of Gaming*. Durham: Routledge.
- Young, G., and Whitty, M. T. (2012). *Transcending Taboos: A Moral and Psychological Examination of Cyberspace*. Durham: Routledge.
- Young, L. (1976). Local Safety Authorities Denounce Game. *The Spokesman-Review*, December 29. Available at: <https://news.google.com/newspapers?nid=1314&dat=19761229&id=eiFOAAAAIIBAJ&sjid=q-0DAAAAIIBAJ&pg=6835,5559430&hl=en> [Accessed August 4, 2015]

Social Networking

The Dialectics of Sharing

Yoni Van Den Eede and Katleen Gabriels

Prologue

In our times of intense use of social networking sites (SNSs), the expression “sharing is caring” has become a veritable motto—used sometimes in earnest, sometimes jokingly, sometimes cynically. Dave Eggers in his novel *The Circle* (2014) takes it a step further. He sketches a picture of a world in which sharing does not just constitute caring: it becomes the norm. In the story, a particularly popular and successful social network company, The Circle, strives toward market dominance, aiming to infect the whole planet with its ideology. That ideology centers on one notion: transparency. The type of transparency that The Circle envisions, however, is radical: it entails transparency of *everything*. As one company slogan reads: “All that happens must be known” (67). In a speech to the employees, one of the founders, Bailey, describes—loudly cheered on by the crowd—the ultimate ideal:

Folks, we’re at the dawn of the Second Enlightenment. . . . an era where we don’t allow the majority of human thought and action and achievement and learning to escape as if from a leaky bucket. We did that once before. It was called the Middle Ages, the Dark Ages. . . . As we all know here at the Circle, transparency leads to peace of mind. . . . ultimate transparency. No filter. See everything. Always. . . . We will become all-seeing, all-knowing.

(67–70)

Eggers’s book reads as a warning cry against the market-imperialist tendencies of currently dominant information and communication technology (ICT) companies such as Facebook and Google. Those seem to be grounded just as much in an ideology of progress and technical rationality: if only we have the right technological means, the right data, the right algorithms at our disposition, combined with the proper mindset of openness and willingness to change, we will be able to create a new and better world.

In themselves, these ideas of course strike a sympathetic chord with many people. To urge for more transparency in political decision making, economy and finance, urban planning, and so on, makes for a commendable aspiration. Sharing—not just in the sense of exchanging information online, also in terms of distributing resources in less wasteful ways (e.g., car sharing)—may just as well *be* caring. Nevertheless, in a not too-far-removed past, sharing was done in a direct way, or at least organized on a fairly overseeable scale: among family members and friends in a private setting, on the level of national social security systems, and globally perhaps in the form of development aid. Today, there is a lot of ‘sharing’ going on, but many of the processes accompanying it plainly escape us. Ironically, sharing in the spirit of transparency elicits a good amount of opacity.

Introduction: Beyond Innocence

SNSs are part and parcel of the lives of many: almost one third of the world population uses an SNS (Kemp 2014). This sounds grand, but in macroscopic historical terms it should not impress terrifically. Older technologies such as agricultural techniques or writing are surely shared by a much greater number of people. Still, SNSs have a different nature than those, if only to the extent that (1) they connect people on a social plane and (2) by them and through them, gigantic databases on human behavior are being produced now, day in, day out.

Human beings are social animals: we exhibit the natural inclination to form groups, networks, communities, and organizations. SNSs by sheer implication of their name—and their predecessors since the wider dispersal of the internet throughout the 1990s, ‘virtual communities’—boost or enhance this process by way of digital ICT. Definitions of SNSs are often formulated along those lines, outlining in the first instance their functionality from the user’s point of view. In a landmark 2007 paper, boyd and Ellison offered up the following, still largely accurate description: SNSs are

web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.

(2007: 211)

When, in those early days, SNSs were just emerging and starting to garner academic interest, they were basically taken at face value. Scholars looked at them from the perspective of what they were supposed to do. Users, developers, marketeers, and researchers alike essentially converged on this point. That is, SNSs purport to support and enhance our natural, unmediated ‘social networking’ activities and dispositions through a technological medium. This did not preclude critical assessments. From the very beginning, as often happens with new technologies (Sturken et al. 2004), in a moral-evaluative sense two camps arose. The ‘utopian’ camp welcomed and lauded SNSs’ potentialities for democratization and emancipation (cf. Tapscott 2008; Shirky 2008). The ‘dystopian’ camp taxed and feared their possible negative effects on our cognitive and social abilities, our culture, our sense of reality (Bauerlein 2008; Siegel 2008; Jackson 2008). Some authors were (already) alarmed by the potential misuse of personal information for corporate or state interests (Fuchs 2008; Hassan 2008; Lyon 2007). But overall—perhaps instigated by the novelty of the phenomenon—everyone’s focus was on the experience of using technology in order to ‘be social.’ Both camps tended to start from the functional viewpoint, asking first and foremost: what do we do with them?

In the post-Snowden era the main question appears to have become: what do *they* do with them? Edward Snowden revealed in 2013, through leaking classified information, the massive online surveillance practices undertaken by the National Security Agency (NSA). Since, we have discovered, quite painfully, how behind the veil of our everyday usage of social media, a whole machinery is hiding. Of course we surmised as much: something must produce the things we see and manipulate on the screen. But this ‘something’ turns out to be much bigger, consequential, and far-reaching than thought. A larger-than-life mirror world—a planet-wide replica of our comings and goings, thoughts, beliefs, intentions—is being created constantly behind the scene, or screen. Within this elusive dataverse, governments and corporations are very actively scurrying around in search of intel and profit.

So the general feeling is different. We have lost some of our naiveté. There is a slight flavor of disenchantment. Philosophical research on social media has yet fully to come to grips with

this vague feeling of a ‘loss of innocence’ of sorts. At the same time many of us are still sharing data on these networks, and clinging heartily to our social media profiles instead of ‘leaving’ SNSs, if only because ‘all our friends are there.’ Looking at how we precisely go about our doings online, from a functional (or phenomenological) perspective, there is not much difference with the ways of old communication—we still engage with others in our close and wider social circles. Yet we know all-too-well, albeit half-consciously, half-unconsciously, that we are sharing everything at the same time with this obscure mirror world, with a system that escapes our attention for a good part.

This chapter explores these still-emergent dialectics. It does so by discussing and problematizing, from the perspective of those ‘dialectics of sharing,’ four conceptual areas that have been fundamental in the philosophical study of SNSs and related phenomena: (1) sociality, (2) self, (3) system, (4) transparency. As will become clear, the first two are to be categorized under the ‘functionalist’ point of view, whereas the latter two already more evoke the flavor of ‘disenchantment’ just sketched.

Sociality: Face-to-Face Versus Mediated

Starting again at the functional end of the spectrum: SNSs serve to help us connect socially. The ‘networked’ sociality that SNSs offer fulfills our strong need to be part of a group (Aronson 1999; Haidt 2003). This certainly goes a long way in explaining their worldwide popularity. Community is an essential human value, since it is the setting “in which a large share of human development occurs” (Feenberg and Bakardjieva 2004: 2).

Constructing communities around media is not new. Humans have long cultivated a sense of community through mediated forms of communication, such as radio and television among others (e.g., Jankowski 2002; Wagman 2012). Since the appearance of computer-mediated virtuality in the 1980s and 1990s, people have been building virtual communities in all shapes. Although such online communities root fundamentally in basic social needs of their users and developers, they display particular characteristics and dynamics (Turkle 1995; Feenberg and Barney 2004).

Our times of broadband and mobile technologies have come to be characterized by a culture of constant connectedness “where perspectives, expressions, experiences and productions are increasingly mediated by social media sites” (van Dijck 2011: 402; also van Dijck 2013). Thanks to the miniaturization and affordability of ICTs such as the smartphone, the distinction between ‘online’ and ‘offline’ has blurred. A key aspect of this hyperconnected culture is that it makes possible instant, real-time interaction between ‘mediated’ selves. In such an environment of immediate, continuous presence, once again new forms of networked sociality arise. The self (see the following discussion) is now ‘always on.’ One’s ‘friends’ are at all times ‘proximal’ in real time, at least potentially.

A recurring issue that has been the subject of much debate since the advent of online sociality concerns the status and value of this mediated or networked sociality. Albert Borgmann (1992, 1999) holds the view, in line with Jean Baudrillard (1995/2008), that online interactions diminish our contact not only with tangible reality but also with others. Due to the lack of bodily intimacy, the mediated encounter can never be ethically meaningful. In contrast to face-to-face interaction, the online self can choose with whom they want to communicate—it is easy to ‘delete,’ ban, or block another person—and in this way feel less responsible for the other. In actual encounters in the flesh, “[t]here is a symmetry between the depth of the world and our bodily incursion into it,” whereas online this symmetry “falls to the level of a shallow if glamorous world and a hyperinformed yet disembodied person” (Borgmann 1992:106).

Recently, these somewhat gloomy pronouncements have been offset by a body of theory and empirical research that points out the tight interwovenness of offline and online sociality. In contrast to other forms of online sociality, such as gaming, virtual worlds, and so on, SNSs are generally based on a system “whereby you befriend people you have pre-existing connections with” (Munn 2012: 5). The virtual Other is often already known. Consequently, networked selves make constant transitions between mediated and nonmediated realms of communication. Seen from these perspectives, ‘continuous connectedness’ is not reductive or extraordinary. On the contrary, it takes place without demanding constant attention. Exactly because online platforms slip into everyday sociality almost unnoticed, we do not experience them as a radical rupture; we naturally integrate them. Especially young people—the so-called digital natives—grow up within this wired sociality, and hence do not question it. Studies show that SNSs strengthen existing relations and friendships (e.g., Ellison et al. 2007; boyd 2008). Networked selves share information and get informed by others. Every ‘like’ or ‘tag’ by our peers has the potential to feel like a warm blanket of reward and self-confirmation (Tamir and Mitchell 2012). Friendship is a virtue and an SNS holds the promise of friendship (Briggle 2008).

The tone of ‘first-generation’ critics such as Borgmann and Baudrillard is still echoed by present-day critics such as Andrew Keen (2008) and Evgeny Morozov (2013). Their pleas to not expect all societal salvation from online interaction are first and foremost ethical in spirit. Also Sherry Turkle (2011), one of the pioneers of early virtual community research, has developed in recent years a more critical stance, being concerned about the deleterious effects of social media and mobile ICTs on face-to-face social skills. These reservations are partly convergent with those of other philosophers such as Hubert Dreyfus (2009) and Don Ihde (2002), who point out on phenomenological grounds that intersubjectivity, identity, and community are intrinsically local and embodied phenomena, based on mutual commitment and dedication. Virtual experience thus always fails to grasp “the full and ordinary richness of bodily experience” (Ihde 2002: 128).

Self: Private Versus Public

Similar issues play out in relation to the next conceptual angle: self and identity. Due to online social networks, our lives have become much more public. The self can be said to be in a state of transition, now that the distances between individuals are immensely decreasing (e.g., Rettberg 2014). This raises the question as to whether one can still speak of a private self, which is of course intrinsically linked to the debate on SNSs and privacy. Identity and the self are no longer ‘isolated’ or ‘private’ in the traditional sense. They become smeared out across a network (Ess 2010).

In a culture of connectivity and immediacy, information no longer travels in closed-off domains but “along a network. Engagement is intrinsically public, taking place within a circuit of larger connectedness” (Birkerts 1994/2010: 494). Selves share personal information on SNSs and, unavoidably, they share private information with unknown others (and with companies and governments). What used to be the exclusive reign of a hypothetical Big Brother has become an activity we all take part in, by shamelessly looking into and commenting upon each other’s lives.

Here, too, research has gone through several phases. Until well into the 1990s, scholars generally assumed the virtual self to be radically detached from the body and the self’s offline characteristics, hence implicitly fostering a Cartesian dualism (e.g., Benedikt 1991). Studies have since convincingly shown that users themselves do not experience such a sharp distinction between the online and the offline (e.g., Markham 1998). On the contrary, online experiences are part of one continuous, experiential reality. They do not take place in a walled-off realm.

This turn in conceptual thought, accompanied “by a (re)turn to embodiment as crucial to how we know the world and thereby what sorts of selves we are” (Ess 2010: 106), has led to new notions of the self and of identity. Conceptualizations such as the “networked self” (Wellman and Haythornthwaite 2002; also Cohen 2012), the “onlife self” (Floridi 2015), and the “smeared-out self” (Ess 2010) emphasize that the self, unlike in atomic and Cartesian views, is distributed across multiple communication networks that “represent hundreds, if not thousands, of simultaneous but potential relationships/engagements” (Ess 2010: 111). Also, it must be added, in contrast to earlier forms of online interactions, SNS users normally post information under their actual names, and no longer anonymously or pseudonymously. This demonstrates the point of online experiences being part of one experienced reality even more: virtual and actual (offline) interactions coevolve and shape one another.

It may thus not surprise that negative experiences on and with SNSs can profoundly affect the self and one’s identity. The Sacco case serves as a painful illustration. The American Justine Sacco intended to make fun of stereotypical racial ideas of WASPs when she sent out the tweet “Going to Africa. Hope I don’t get AIDS. Just kidding. I’m White!” on December 20, 2013. The tweet went viral and affected her social life deeply (obviously, an actual world effect). She got pilloried, lost her job, and—as it is impossible to delete this information from the internet—the 140 characters have haunted her since, making it difficult if not impossible for her to find new employment. This is a strong example of the extent to which our lives have become networked and, more practically, of how cautious one has to be with online posts. Many do seem to mistake SNS updates for private speech and communication. It may also be a sign of a certain culture of mass, instant, and exaggerated moral indignation, but this is a discussion in itself.

At the same time, while there is a decrease of the private self and private space, the increase of the narcissistic self—apparently paradoxically—cannot be disregarded either (Bakardjieva and Gaden 2012). SNSs are highly self-focused, since they are designed as personal networks where “[e]ach user is at the center of their community” (La Barbera, La Paglia, and Valsavoia 2009: 34). It appears that about 80% of the content posted on SNSs is about people’s subjective and immediate experiences (Naaman et al. 2010). One may also think of the selfie (and selfie stick) phenomenon. Importantly, this culture of the self surely has an emancipating and liberating side to it, too: citizen journalism, for instance, gives ordinary people a voice, and people can become influential ‘media’ themselves (e.g., through popular Facebook pages).

System: Liberation or Domination

Until now the focus has been mainly on *what we do with* SNSs. Seen from that more ‘instrumental’ perspective, SNSs either enhance or distort our in-bred sociality, and they harbor either an emancipatory or a threatening potential for our identity-building projects. In postphenomenological parlance, they harbor “amplifications” as well as “reductions” (Ihde 1990), at least from the perspective of the individual or of noncommercial social assemblages. However, focusing too much on SNSs’ functional use and on our personal experience in relation to them, we might almost forget that behind their slick interfaces and immediate usefulness, mostly corporations and other commercial entities hide. Those can be expected to have specific agendas: profit making, securing market share, appeasing shareholders, etc. Obviously, these endeavors can conflict with users’ interests. Beyond the user interface lies an algorithmic system that the user only minimally controls.

In recent years, many debates have been waged over issues surrounding liberation and domination. On one side, there is mostly a commercial organization—Facebook, Google, and so on—attempting to find (new) ways to make money through its platform. This may involve showing more advertisements, or selling profile data to third parties. On the other side, there are

users contesting these strategies for reasons of ease of use or, in most cases, privacy. In practice one often observes a ‘trial-and-withdraw’ dynamic on the part of companies. Terms of Service (ToS) are changed in order to enable a certain type of commercial exploitation—for instance the utilization of location data or posted pictures for commercial purposes—where after these changes are denounced by the public, and subsequently (at least partially) rolled back by the company (Silverman 2015). Often, the proposed changes are so opaquely worded that it takes a substantial amount of interpretation before any critique can be launched. Nonetheless, a give-and-take of sorts between corporate and social-individual interests can be discerned.

Depending on which pole of that spectrum between ‘domination’ and ‘liberation’ one wishes to emphasize, a different story about SNS’s political-economic underpinnings emerges. Stressing domination casts users as unsuspecting or foolish pawns in the hands of powerful state or corporate organisms. They may be the guileless labor force in “semicapitalism” (Berardi 2009) and “cognitive capitalism” (Moulier-Boutang 2011). Others perceive them as subjects of the “performance society,” that differs from the previous, “command society,” famously analyzed by Foucault, in that our current situation entails we are no longer surveilled and disciplined in a top-down way: we surveil and discipline ourselves and each other (Han 2010; Han 2012). Emphasizing liberation, by contrast, elicits so-called success stories in the style of the ‘Twitter Revolution’ rhetoric that hailed SNSs as being instrumental in bringing about, amongst others, the Arab Spring uprisings in Tunisia, Libya, and Egypt, and as helpful in enabling and constituting a more democratic public sphere. In between these two extremes—that in actuality almost never appear in their pure form—the most diverse assessments can be found that see liberatory as well as disciplinary possibilities in SNSs and digital network technologies. Many of these, though not all, are Critical Theory-inspired (Fuchs 2014; Lovink 2011; Castells 2013; see also, specifically in relation to the Arab Spring, Howard and Hussain 2013).

One prominent approach in this regard (combining both poles), which in fact assesses how the two are exactly intertwined, is philosopher Andrew Feenberg’s Critical Theory of Technology (2002). In a first movement, in his view, technology gets deployed to the advantage of an elite, consolidating its power. He calls this the preservation of “operational autonomy.” Yet, in a second movement, options remain for users, consumers, and interest groups to question, modify, and reject technologies. Feenberg terms this process “subversive rationalization” (1995). When looking at a technology, thus, we must consider its potentials and not repudiate it straight-out, out of fear or worry. With regard to the internet, it is simply too early to evaluate its value for democracy; he points out: it is still a developing technology, and so we cannot judge it thoroughly yet (Feenberg 2009: 77). But, we must also not shove it aside as “endless talk” “discussion lies at the heart of a democratic polity. Any new scene on which it unfolds enhances the public sphere” (ibid., p. 78). The fact remains that, given the “first movement” described earlier, powerful actors will at all times seek to steer its development toward their own ends. It is in this context that our evaluations must find their starting point, and so vigilance is required, that can concretely take shape by way of what Feenberg calls “technical micropolitics,” comprising the demanding of design changes, participatory design, creative appropriation, and legislative action.

Transparency: From Sharing to Algorithmic Culture

The question raised by social networking is not a question of a guileless social self pitted against a dark, faceless system. The dialectics of sharing are more complex than that. The ‘behind the scenes’ cannot be separated from what takes place ‘on the scene’ (i.e., social experiences and self-profiling). An essential concept in this regard is transparency. Unlike in *The Circle*, this does not necessarily have to take dystopian forms. We appreciate throughout our

everyday social interactions some amount of transparency, and we (mistakenly?) expect it to a high degree from administrations and bureaucracies, and from governmental and corporate decision making. The same counts for sharing: it is a ‘natural’ thing to share, and so it feels just as ‘natural’ to use an SNS to engage in sharing. The emergence of sharing economies over the last couple of years—e.g., car sharing, property sharing, crowdfunding, virtual currencies, etc.—is partly grounded in SNS-type technologies: digital networks functioning like or even on the basis of existing platforms such as Facebook. More transparency and more sharing, considered as abstract phenomena, are experienced in the first instance as kinds of progress, in line with Enlightenment ideals (Vattimo 1992).

At what point exactly these benefits turn into disadvantages, or even dangers, remains a matter of contention. As illustrated perhaps grotesquely—but still—in *The Circle*, somewhere along the line, a legitimate desire for transparency and sharing threatens to turn into a controversial demand for them. Put simplistically—but still—this usually happens when some company sees opportunities for profit making, or some government sees possibilities for surveillance. The problem, however, gets exacerbated because hardly any clear line of demarcation can be drawn between our ‘natural’ dispositions for sharing and openness and their more imposed, tyrannical, asphyxiating versions (see also Van Den Eede 2010). Thus a ‘culture’ of transparency and openness—epitomized in the phrase “I/we have got nothing to hide”—arises, in which openness of everyone and everything is required *and* praised, from ‘open’ offices and workplaces to ‘open’ kitchens in restaurants; in which many people even gladly and avidly share their sports activities online, how many miles they ran and calories they burned, how much weight they lost, and so on; in which employers are even suspicious of an aspiring employee if the latter does not have a social media profile. Parts of our personal lives have become *intimate capital* (a term hailing from Dutch philosopher Stine Jensen, 2011) to be staked out in the market of not just professional, but all relationships. Clearly, it is not only governments and corporations that demand sharing and transparency, *we* to a large extent demand it *from each other*: spouse from spouse, friend from friend, parent from child, etc.

Our times have been framed as a ‘culture of connectivity.’ Our shared desire for transparency—both bottom-up and top-down—has also engendered a data-driven ‘algorithmic culture,’ that is, a society in which we have come to approach and understand ourselves, others, and our environment increasingly through algorithms. In Pasquale’s phrase: “[a]s we are treated algorithmically (i.e., as a set of data points subject to pattern recognition engines), we are conditioned to treat others similarly” (2015, n.p.). Developers motivate people to participate in this algorithmic culture, and companies subsequently gain profound insight in users’ behaviors thanks to the gigantic databases that are being produced this way. The algorithmic culture is driven by economic and commercial logics, and treats data as a new currency. Whereas in the past it was difficult to obtain reliable and precise information, the internet gives data brokers (or data controllers) a detailed front-row view of behavioral patterns and predictive analytics, which are, among other things, interesting for personalized marketing objectives.

Life has become increasingly quantified. It may not surprise that companies, from an economic viewpoint, aspire to connect as many people as possible in this algorithmic culture. What is perhaps more surprising is that people are so easily convinced to promote their algorithmic selves online. Apps can help them find a restaurant, parking spot, partner (e.g., on Tinder), so they benefit from it as well and willingly cooperate. For instance, whereas running is in practice often a solitary activity, athletic activity services such as Strava or RunKeeper operate via internet-connected devices such as smartphones; physical performances are subsequently logged and communicated through SNSs with others, who can compare the graphs and statistics with their own performances. Such people who share data, logged from their bodily activities (e.g., miles they ran, calories they burned, and so on), act in the spirit of the “quantified self

movement” (Ruckenstein 2014; Swan 2013). A hyperconnected self always leaves a digital trace.

An algorithmic culture, putting so much emphasis on data, leaves little room for experiences and realities that cannot be easily measured or quantified, but that are nonetheless intrinsically meaningful and essential. Several theorists have expressed the worry that we face “a disenchanting world, a dead world-machine with only quantities, no qualities, facts with no value, and mechanisms without purpose” (Doede 2009: 52; also Morozov 2013; Carr 2014). These critical voices share the concern that we have become overly dependent on computers, automation, and efficiency. Even though we must not lose sight of the positive aspects, it is essential to underscore the importance of increasing and stimulating critical awareness and digital literacy (and user empowerment) throughout research and education. Not only does the public and philosophical debate need to be attentive to “the good life” (Higgs et al. 2000; Brey et al. 2014), it also needs to take stock of how this very notion changes due to technological developments and unethical practices such as secretly selling, controlling, and manipulating data. This way, the notion of transparency remains most pertinent, since—ironically—companies and governments are insufficiently transparent toward consumers and citizens.

Conclusion

This chapter sought to offer an outlook on the views and concepts at stake in the context of what it described as the emergent *dialectics of sharing*, typifying our hyperconnected culture. The rise and spread of SNSs have affected society and culture both rapidly and deeply. Early philosophical and other research of SNSs took them mostly at face value. Stern criticisms were already developed, but framed still first and foremost in terms of the functional aspects of the technologies under scrutiny. Since then, scholarship has been progressing past those first approaches. After Edward Snowden’s revelations, our understanding of social media lost its ‘functionalist’ innocence to a good extent.

Four interrelated conceptual areas were elaborated on: sociality, self, system, and transparency. Four fields of tension in particular traverse these conceptual angles and were prominent throughout the chapter. Some takeaway points can be formulated with regard to all. First, the dichotomy between face-to-face and mediated: the difference between online and offline modalities has blurred substantially, though community can still be said to root essentially in local, embodied experience.

Second, private vs. public: notwithstanding the fact that all the information shared on an SNS is essentially public (shared on and with a network), many still mistake it for private communication and consequently share intimate details not intended for a wide audience. This has led to an interesting ‘clash’ concerning selfhood, as there is a decrease of the private self, and yet at the same time an increase of the ‘narcissistic’ self cannot be ignored.

Third, the tension between liberation and domination: although the internet was hailed as a democratizing force throughout the 1980s and 1990s, it has become highly dominated by commercial interests. Due to this market dominance, SNS’s potentially liberating effects are in sharp contrast with its possible (or likely?) alienating effects, which raises the compelling question as to what extent users can still be free and autonomous: if they are constantly being watched, if their data are stored in databases, and if these data are analyzed for commercial and surveillance-oriented objectives. In addition, we must not lose sight of the long term, and often unforeseen, effects, as data do not disappear (the ‘right to be forgotten’).

Fourth, synthesizing the previous three: the dialectics of sharing. Everything we share as individual users on SNSs is shared at the same time with governments, companies, ‘invisible’ and unknown others, and so on. Users are forced to navigate this interwovenness: even

though they know the trade-offs and are informed, they will not easily withdraw or cancel their accounts, exactly because their (personal, social) network is represented online, so it ultimately becomes a trade-off between privacy and sociality.

What direction should philosophy of technology take with regard to these themes? The topics are still relatively understudied within the field (not counting a few notable exceptions, e.g., Wittkower 2010). Judging from this overview of dichotomies, one can expect the discipline to be extremely well-suited to a comprehensive and critical approach to SNSs. Contemporary philosophy of technology is characterized by a sensitivity for perspectives that go beyond either instrumentalist-functionalist or reifying views. One may think particularly of accounts of “mediation” (Verbeek 2005; Verbeek 2006), human-technology relations (Ihde 1990; Rosenberger and Verbeek 2015), and “relationality” (Sharon 2014; Coeckelbergh 2013). Instead of self-sufficient entities and ‘things,’ these approaches see interrelationships. A fertile prospect for the philosophical study of SNS would be to engage more with these frameworks, in order to sharply draw further, and better understand, the dialectics of sharing. The most important task for philosophy of technology may consist in making ‘transparency’s opacity’ sufficiently and comprehensibly . . . transparent.

References

- Aronson, E. (1999). *The Social Animal*, 8th edition. New York: Worth Publishers.
- Bakardjieva, M., and Gaden, G. (2012). Web 2.0 Technologies of the Self. *Philosophy & Technology*, 25(3), 399–413.
- Baudrillard, J. (1995/2008 [original 1995: *Le Crime Parfait*]). *The Perfect Crime*. London: Verso.
- Bauerlein, M. (2008). *The Dumbest Generation: How the Digital Age Stupefies Young Americans and Jeopardizes Our Future*. New York: Jeremy P. Tarcher.
- Benedikt, M. (ed.) (1991). *Cyberspace: First Steps*. Cambridge, MA: MIT Press.
- Berardi, F. (2009). *Precarious Rhapsody: Semiocapitalism and the Pathologies of the Post-Alpha Generation*. London: Minor Compositions.
- Birkerts, S. (1994/2010). Into the Electronic Millennium. In C. Hanks (ed.) *Technology and Values: Essential Readings*. Oxford: Wiley-Blackwell, 491–499. Original: Birkerts, S. (1994). *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*. New York: Fawcett Columbine, 117–133.
- Borgmann, A. (1992). *Crossing the Postmodern Divide*. Chicago: The University of Chicago Press.
- Borgmann, A. (1999). *Holding on to Reality: The Nature of Information at the Turn of the Millennium*. Chicago: The University of Chicago Press.
- boyd, d. (2008). Why Youth (Heart) Social Networks: The Role of Networked Publics in Teenage Social Life. In D. Buckingham (ed.) *Youth, Identity, and Digital Media*. Cambridge, MA: MIT Press, 119–142.
- boyd, d., and Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13, 210–230.
- Brey, P., Briggie, A., and Spence, E. (eds.) (2014). *The Good Life in a Technological Age*. London: Routledge.
- Briggie, A. (2008). Real Friends: How the Internet Can Foster Friendship. *Ethics and Information Technology*, 10(1), 71–79.
- Carr, N. (2014). *The Glass Cage: Automation and Us*. New York: W. W. Norton & Company.
- Castells, M. (2013). *Communication Power*, 2nd edition. Oxford: Oxford University Press.
- Coeckelbergh, M. (2013). *Human Being @ Risk: Enhancement, Technology, and the Evaluation of Vulnerability Transformations*. Dordrecht: Springer.
- Cohen, J. E. (2012). *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*. New Haven, CT: Yale University Press.
- Doede, B. (2009). Transhumanism, Technology, and the Future: Posthumanity Emerging or Sub-Humanity Descending? *Appraisal*, 7(3), 39–54.
- Dreyfus, H. L. (2009). *On the Internet*, Revised 2nd edition. London: Routledge.
- Eggers, D. (2014). *The Circle*. London: Penguin.

- Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The Benefits of Facebook 'Friends': Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, 12, 1143–1168.
- Ess, C. (2010). The Embodied Self in a Digital Age: Possibilities, Risks and Prospects for a Pluralistic (Democratic/Liberal) Future? *Nordicom Information*, 32(2), 105–118.
- Feenberg, A. (1995). Subversive Rationalization: Technology, Power, and Democracy. In A. Feenberg and A. Hannay (eds.) *Technology and the Politics of Knowledge*. Bloomington, IN: Indiana University Press, 3–22.
- Feenberg, A. (2002). *Transforming Technology: A Critical Theory Revisited*. Oxford: Oxford University Press.
- Feenberg, A. (2009). Critical Theory of Communication Technology: Introduction to the Special Section. *The Information Society*, 25(2), 77–83.
- Feenberg, A., and Bakardjieva, M. (2004). Consumers or Citizens? The Online Community Debate. In A. Feenberg and D. Barney (eds.) *Community in the Digital Age: Philosophy and Practice*. Lanham: Rowman & Littlefield Publishers, 1–28.
- Feenberg, A., and Barney, D. (eds.) (2004). *Community in the Digital Age: Philosophy and Practice*. Lanham: Rowman & Littlefield Publishers.
- Floridi, L. (ed.) (2015). *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Cham: Springer.
- Fuchs, C. (2008). Review of 'Wikinomics: How Mass Collaboration Changes Everything,' by Don Tapscott & Anthony D. Williams. *International Journal of Communication*, 2, 1–11.
- Fuchs, C. (2014). *Social Media: A Critical Introduction*. London: Sage.
- Haidt, J. (2003). The Moral Emotions. In R. J. Davidson, K. R. Schere, and H. H. Goldsmith (eds.) *Handbook of Affective Sciences*. Oxford: Oxford University Press, 852–870.
- Han, B.-C. (2010). *Müdigkeitsgesellschaft*. Berlin: Matthes & Seitz.
- Han, B.-C. (2012). *Transparenzgesellschaft*. Berlin: Matthes & Seitz.
- Hassan, R. (2008). *The Information Society*. Cambridge: Polity.
- Higgs, E., Light, A., and Strong, D. (eds.) (2000). *Technology and the Good Life?* Chicago: The University of Chicago Press.
- Howard, P. N., and Hussain, M. M. (2013). *Democracy's Fourth Wave? Digital Media and the Arab Spring (Oxford Studies in Digital Politics)*. New York: Oxford University Press.
- Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth*. Bloomington, IN: Indiana University Press.
- Ihde, D. (2002). *Bodies in Technology*. Minneapolis, MN: University of Minnesota Press.
- Jackson, M. (2008). *Distracted: The Erosion of Attention and the Coming Dark Age*. Amherst, NY: Prometheus Books.
- Jankowski, N. W. (2002). Creating Community with Media: History, Theories and Scientific Investigations. In L. A. Lievrouw and S. Livingstone (eds.) *Handbook of New Media: Social Shaping and Consequences of ICTs*. London: Sage, 34–49.
- Jensen, S. (2011). *Echte vrienden: Intimiteit in tijden van Facebook, GeenStijl en WikiLeaks*. Rotterdam: Lemniscaat.
- Keen, A. (2008). *The Cult of the Amateur: How Blogs, MySpace, YouTube and the Rest of Today's User-Generated Media Are Killing Our Culture and Economy*. London: Nicholas Brealey Publishing.
- Kemp, S. (2014). Global Social Media Users Pass 2 Billion. *We Are Social*. Available at: <http://weare-social.net/blog/2014/08/global-social-media-users-pass-2-billion/> [Accessed May 27, 2015].
- La Barbera, D., La Paglia, F., and Valsavoia, R. (2009). Social Network and Addiction. In B. K. Wiederhold and G. Riva (eds.) *Annual Review of Cybertherapy and Telemedicine*, vol. 7. Amsterdam: IOS Press, 33–36.
- Lovink, G. (2011). *Networks Without a Cause: A Critique of Social Media*. Cambridge: Polity.
- Lyon, D. (2007). *Surveillance Studies: An Overview*. Cambridge: Polity.
- Markham, A. N. (1998). *Life Online: Researching Real Experience in Virtual Space*. Walnut Creek: AltaMira Press.
- Morozov, E. (2013). *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist*. London: Allen Lane.

- Moulier-Boutang, Y. (2011). *Cognitive Capitalism*. Cambridge: Polity.
- Munn, N. J. (2012). The Reality of Friendship Within Immersive Virtual Worlds. *Ethics and Information Technology*, 14, 1–10.
- Naaman, M., Boase, J., and Lai, C. H. (2010). Is It Really About Me? Message Content in Social Awareness Streams. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Association for Computing Machinery), Savannah, GA, 189–192.
- Pasquale, F. (2015). The Algorithmic Self. *The Hedgehog Review: Critical Reflections on Contemporary Culture*, 17(1). Available at: www.iasc-culture.org/THR/THR_article_2015_Spring_Pasquale.php [Accessed May 29, 2015].
- Rettberg, J. W. (2014). *Seeing Ourselves through Technology: How We Use Selfies, Blogs and Wearable Devices to See and Shape Ourselves*. Basingstoke: Palgrave Macmillan.
- Rosenberger, R., and Verbeek, P.-P. (eds.) (2015). *Postphenomenological Investigations: Essays on Human—Technology Relations*. Lanham: Lexington Books.
- Ruckenstein, M. (2014). Visualized and Interacted Life: Personal Analytics and Engagements With Data Doubles. *Societies*, 4(1), 68–84.
- Sharon, T. (2014). *Human Nature in an Age of Biotechnology: The Case for Mediated Posthumanism*. Dordrecht: Springer.
- Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*. London: Allen Lane.
- Siegel, L. (2008). *Against the Machine: Being Human in the Era of the Electronic Mob*. New York: Serpent's Tail.
- Silverman, J. (2015). *Terms of Service: Social Media and the Price of Constant Connection*. New York: Harper.
- Sturken, M., Thomas, D., and Ball-Rokeach, S. J. (2004). *Technological Visions: The Hopes and Fears That Shape New Technologies*. Philadelphia, PA: Temple University Press.
- Swan, M. (2013). The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data*, 1(2), 85–99.
- Tamir, D. I., and Mitchell, J. P. (2012). Disclosing Information About the Self Is Intrinsically Rewarding. *PNAS*, 109(21), 8038–8043.
- Tapscott, D. (2008). *Grown Up Digital: How the Net Generation Is Changing Your World*. New York: McGraw-Hill.
- Turkle, S. (1995). *Life on the Screen: Identity in the Age of the Internet*. New York: Simon & Schuster.
- Turkle, S. (2011). *Alone Together: Why We Expect More From Technology and Less From Each Other*. New York: Basic Books.
- Van Den Eede, Y. (2010). Collecting Our Lives Online: The Use of Digital Media Seen Through the Lens of Collecting Practices. *Techné: Research in Philosophy and Technology*, 14(2), 103–123.
- van Dijck, J. (2011). Flickr and the Culture of Connectivity: Sharing Views, Experiences, Memories. *Memory Studies*, 4(4), 401–415.
- van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*. Oxford: Oxford University Press.
- Vattimo, G. (1992). *The Transparent Society*. Cambridge: Polity Press.
- Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: The Pennsylvania State University Press.
- Verbeek, P.-P. (2006). Acting Artifacts: The Technological Mediation of Action. In P.-P. Verbeek and A. Slob (eds.) *User Behavior and Technology Development: Shaping Sustainable Relations Between Consumers and Technologies*. Dordrecht: Springer, 53–60.
- Wagman, I. (2012). Télé-Clubs and European Television History Beyond the Screen. *VIEW—Journal of European Television History and Culture*, 1. Available at: www.viewjournal.eu/index.php/view/article/view/jethc024/46 [Accessed May 29, 2015].
- Wellman, B., and Haythornthwaite, C. (eds.) (2002). *The Internet in Everyday Life*. Oxford: Blackwell.
- Wittkower, D. E. (ed.) (2010). *Facebook and Philosophy: What's on Your Mind?* La Salle, IL: Open Court Press.

Further Readings

- Ellison, N. B., and boyd, d. (2013). Sociality Through Social Network Sites. In W. H. Dutton (ed.) *The Oxford Handbook of Internet Studies*. Oxford: Oxford University Press, 151–172. Book chapter that provides, among other things, a well-grounded definition of SNSs, their origins and evolution, and research challenges.
- Ess, C. (2012). *Digital Media Ethics*. Cambridge: Polity Press. Introductory monograph on ethics and digital media.
- Papacharissi, Z. (ed.) (2010). *A Networked Self: Identity, Community, and Culture on Social Network Sites*. New York: Taylor & Francis. Edited volume on self-presentation and construction of the self on SNSs.
- Vallor, S. (2012). Social Networking and Ethics. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*, Winter ed., Available at: <http://plato.stanford.edu/archives/win2012/entries/ethics-social-net-working/>. Comprehensive online article that addresses the ethical implications of SNSs.

Special issues of Ethics and Information Technology

- (2010). *Ethics and Information Technology 12*, eds. M. C. Bottis, F. C. Grodzinsky, and H. T. Tavani. Special issue on moral luck, social networking sites, and trust on the web.
- (2012). *Ethics and Information Technology 14*, eds. M. Cocking, J. van den Hoven, and J. Timmermans. Special issue on online friendship.

Technoart

Don Ihde

What some have called the *praxis* turn in science studies, during which attention was focused upon laboratories and practices rather than theory production, led to the recognition that science proceeds by using technologies (instruments), which in turn led to the term *technoscience*. Here I will use the same methodological features to examine a deep history and phenomenology of art praxis in which a parallel result leads to my neologism, *technoart*. I take my clues from what today is called *postphenomenology*, which emphasizes human relations to technologies, the resultant forms of embodiment skills developed in uses, and the multistable results over time.

My tactic will follow three unequal time slices in human art history: What I shall call *Paleolithic and Ice Age Art* actually goes back to beginnings –100,000 years ago through to the end of the Ice Ages, –10,000. (I employ a science convention that sometimes uses a year, BP or “before present” or, more recently, simply a “–” sign before the year. This convention avoids BD, AD or BCE, CE datings. I shall revert to AD and CE datings for historic times.) There are two reasons for the first slice being so long (–100,000 > –10,000). The first is that during this period of time the tool kits or technologies change little. My favorite long-term technology is the Acheulean hand axe, which goes back 1.8 million years and undergoes very little change until it was abandoned –400,000 years ago. The second reason relates to an equally slow change in what broadly could be called an ‘art style.’

The second slice, which roughly coincides with the European Renaissance, fourteenth through seventeenth centuries (AD or CE, our time), is much shorter but encompasses a period of vast technological experimentation that used a range of tools such as optics, ranging through a proliferation of musical instruments, many of which were used in both art and science practices.

The third slice is narrower still. I shall turn to mostly digital art technologies, which are largely twentieth through twenty-first century (AD or CE) inventions and for which change is now so rapid that it is difficult to stay abreast of the latest innovations. As with previous slices, I shall look at both visual and auditory—and other arts—in the examples.

Finally, before leaving my introductory comments, I will insert here an anecdotal bit about interests and qualifications related in part to the special role played by phenomenology or post-phenomenology, which always includes the role of human experience. Art and science—better, technoart and technoscience—have been core interests of mine back to boyhood. During and a bit beyond my graduate school experiences, I was on the side a painter, roughly late 1950s through 1970. Then, there followed a long fallow period of no painting, rather books (many of which did include my own drawings and illustrations), until 2008 when I resumed painting. Today I continue art practices in my studio in Vermont and regularly exhibit works in galleries and art centers. My collections of art books are substantial and upon retirement to Manhattan, my wife and I regularly attend art exhibits, galleries, and concerts on a weekly basis. Of particular interests have been, on the one hand, music—and readers may know my *Listening and Voice: Phenomenologies of Sound*, now a second edition (2007) and my recent *Acoustic*

Technics (2015). On the other hand, I have long been interested in Paleolithic and Ice Age art, with a library full of books to which much personal observation can be added: for example, cave art (Niaux, France), cliff art (South African bushmen, U.S. Southwest, Mexico), Scandinavian petroglyphs, and Paleolithic grave mounds and monuments (United Kingdom, Scandinavia, Korea, South America, etc.). Indeed, since retirement after forty-plus years of technoscience work, the return to art practice and the comparison of art and science praxis is a major interest.

Slice I: Paleolithic to Ice Age Technoart

According to current theory, modern humans originated $-200,000 > -100,000$ years ago with African origins. With the emergence of *homo sapiens* comes what Jill Cook calls the “modern mind” in *Ice Age Art: Arrival of the Modern Mind* (2013). This book was the catalogue for the largest-ever collection of Paleolithic to Ice Age Art objects for an exhibit. Objects came from all over Europe and some of Eurasia, encompassing art tools, sculptures, decorated weapons. This reference has proved invaluable for the $-40,000 > -10,000$ portion of my slices. Supplemented by my references to Altamira, Chauvet, and Lascaux cave paintings, the best of Eurasian and European paintings discovered to date can be dealt with. Even older African beginnings ($-100,000 > -75,000$) rely upon more scattered references. I begin with these and with an emphasis upon tools: Paleolithic technoart.

Blombos Cave, South Africa, yielded a full pigment-mixing tool kit and was discovered in 2011. (I note that the oldest caves tend to be the most recent discoveries: Niaux, Altamira, and Lascaux were nineteenth-century discoveries; Chauvet, with the oldest European cave art, was discovered in 1995; and Maros, Indonesia was discovered in the early twentieth century, but work on its art begins in 1950. These caves go back to $-40,000 > -39,000$.) In Blombos, giant sea snail shells were used as containers; bones of a seal were used to grind pigment granules of red and yellow ochre; and charcoal combined with a liquid, marrow, to produce the typical black, red, and yellow colors that pervade Paleolithic and Ice Age art from this beginning until the end of the Ice Age in roughly $-10,000$. While no representational paintings are of this age, ochre-stained beads, probably body paint, and burial artifacts have been found, along with cross-hatched shells that can be dated from $-75,000$. In a section to follow, “Image and Logic,” I describe how it appears that the first markings are abstract, geometrical, or of the ‘logic’ style.

The current consensus is that modern humans began to move to and settle in Europe roughly $-45,000$. By at least $-42,000$, Ice Age art proliferates through much of this northern territory. Maros, Indonesia, and possibly some Australian cliff art goes back $-28,000$ as well.

Detour: Niaux Cave, France

During 1967–1968 I was living in Paris, France, doing my first book, *Hermeneutic Phenomenology: The Philosophy of Paul Ricoeur* (1971), which turned out to be the year of the “Events of May,” which closed down virtually everything in France. On the first plane into Orly, my friend Bill Chafe, a historian from the University of North Carolina, and his wife flew in. With my first wife, Carolyn, we rented a car and headed to the south of France to Niaux, an Ice Age cave that had been recommended to me. We arrived and one German couple was there as equally early arrivals. So, with a guide, the seven of us entered the cave. The only lighting was the guide’s flashlight. For over a kilometer we trudged on the damp floors into the inner recesses. Once well beyond any outside light, there were abstract spears drawn on the walls and one could see Paleolithic-era barefoot prints on the floor. Finally, reaching a curve in the cave, we turned and soon walls of animals appeared: bison, horses, ibex, and more. They were shockingly ‘realistic,’ with many having humps and curves utilizing bumps and curves of the cave wall.

These are obviously old, since some are covered deeply with calcite. Then comes a shock: our guide turns off his flashlight—deep dark, deeper than any other kind of dark—and he lights a candle. In the flickering light the animals begin to ‘breathe’ and we are transfixed into an Ice Age time, an experience never before undergone and never to be forgotten. Flickering light and cave wall undulations remain part of this art throughout.

Image and Logic

Peter Galison, masterful historian and philosopher of science, first pointed out in his *How Experiments End* (1987) how microphysics with its complex instruments had “instrumental traditions.” Later, his *Image and Logic* (1997) argued that some physicists favored, on one side, an image tradition, seeing sometimes in a single, gestalt image, a feature of subparticle phenomena, and other physicists followed a logic tradition, favoring counting, multiple building data, and would deride any single image insight. Years after my 1968 experience, I realized that Paleolithic humans, too, had something of an image and logic approach to Ice Age art. Niaux had its abstract spears and pointers, and these exist in many of the Ice Age caves, along with hatch marks, stick figures, and geometrical markings whose meanings are forgotten and unknown. In contrast, the animal, fish, and bird figures are identifiable in spite of both the millennia of their constructions or of the time gap from then to now.

Paleolithic Subslice I: $-40,000 > -20,000$

I now turn to one subslice of Paleolithic art, roughly $-40,000 > -20,000$, then move to a second subslice, $-20,000 > -10,000$. (I note in passing that the dating is dicey. Newer and more accurate Carbon 14 dating, sometimes adumbrated with other methods such as thermo-luminescence, has in most cases pushed back earlier dating, often by thousands of years.) If we assume the convention regarding early *homo sapiens* arrivals in Europe to be roughly $-45,000$, then by $-42,000 > -40,000$ there is a proliferation of Paleolithic art: cave and cliff paintings, sculptures, pendants, decorated objects, musical instruments, evidence of dance, etc. I shall highlight only a few examples with special references to embodiment skills and technoart tools.

Beginnings: $-100,000 > -75,000$

- Pigment kit: a prehistoric ‘chemistry’ mix to attain black, red, and yellow paints; mixing skills
- Shell beads: some colored, tools needed, drill for holes, skills to manufacture
- Unknown other color uses: body paint? Funeral practices? Art objects?

Out of Africa, $-40,000 > -20,000$

- Pigment colors continue: black, red, and yellow (some shading experiments) earliest construction indicates much hand work (palm and finger prints remain on clay works) but many drawings show etching outlines requiring a burin or etching device with attendant drawing skills. Brushes have not been invented, but blowing (with mouth or tubes for stencils and some effects on paintings), sponges, or pads used.
- Scaffolding, some elaborate in large caves for ceiling paintings. Carpentry skills.
- Older images, often monochrome: black dominant, red, many engraved or shaped in clay, some with shading. Earliest parts of chronology lead to later coloring-in with shaded ochres

and black lines, Art skills were high, with very few ‘amateur’ or beginner works found. Hand stencils, though, include adult male and female, adolescent, and even children’s hand stencils (and in one slightly more recent example from Libya, stencils were made of lizard paw prints, –8,000). Note, too, that in the earliest part of this period, mastodons, rhinoceros, aurochs, cave bears, and lions are common—and by the end of this period many have begun to go extinct! Large animals predominate, with some birds and fish, but hares, rabbits, and small animals—which doubtless were part of the diet—are missing.

- Special reference to Maros cave in Indonesia in this period finds the same palette and style, but with different animals, including pig-deer and dwarf buffalos.
- Multiple small statues, dominated by female ‘Venus’ figures, some early, more later in –40,000 > –20,000 period. Also small animal statues of extant animals, many horses, bison, reindeer, lions, etc. Carving skills are related in part to knapping skills for weapons and tools, hand-axes, and artifacts.
- Much decoration, etched on batons, spear-throwers, antlers, and bones. Again, art etching skills. Materials include ivory, bone, stone, antlers, and clay.
- I have emphasized visual two and three-dimensional art, but music (and dance) was also very much part of this technoart scene. A –45,000 four-holed bear bone flute has been found, associated with a Neanderthal site. Bone flutes from the German Hohe-Fels cave, vulture and swan, date to –36,000. (Note: by doing a physics of hole placement it appears that the tuning system fills the upper diatonic scale.) Add dance-pattern barefoot prints from deep caves that indicate dance practices.

What may be seen here is that Paleolithic art was already in full bloom in this –40,000 > –20,000 period. Although its palette and materials were limited, its technologies were already extensive, with many art tools and extending to scaffolding and lighting technologies.

Paleolithic Subslice 2: –20,000 > –10,000

By the time of this period’s closure, the end of the Ice Age is looming, but we are still prehistorically before agriculture and writing. Niaux, which I visited, has its oldest images –14,000, and thus is late in the Paleolithic period. By midpoint, art remains are now global: the Americas and all of Asia are included in technoart production. Jill Cook, in *Ice Age Art*, claims that this period constitutes something of a Paleolithic Renaissance in the proliferation of art and of innovations. Her dating runs from –22,000 > –10,000 and takes note of very public cliff friezes along cliffs down to the ingenious “swimming reindeer” statuette of this period. What I note is possibly a move outside—from caves to cliffs (Cook 2013).¹ I shall remark only on some long-lasting trajectories. If there was an Ice Age Renaissance, it remained continuous with its long prehistory with regard to palette and to some extent instruments. Blues and greens remain absent from the palettes. While there is an etching of a ‘Venus’ holding a horn, actual musical artifacts remain those of the ‘bonewind’ variety. (I do not doubt there were percussion and probably horn instruments.) Clearly there were instrumental innovations spanning art and other daily practices. Many shaped burins have been found for incising and carving; in addition, spears along with spear straighteners, and spear throwers, possibly even archery existed (Cook 2013).² By –12,000 most of the large animals were extinct (cave bears –24,000; rhinos, mammoths, and cave lions, roughly –12,000). Probably most of these were rare to the extreme by –20,000. It does seem that as humans emerged from the long Ice Ages, they were ready for more experimentation, which led, as we know, to the domestications of agriculture with respect to both animals and plants, and later to the practices of inscription in the form of writing (from –10,000 > –7,000).

Slice 2: Renaissance Technoart, Fourteenth < Seventeenth Centuries AD/CE

Lynn White, Jr., in his *Medieval Technology and Social Change* (1962), argued that from late medieval times to the Renaissance, there was a major technological revolution. Machines, some large and complex, began to replace human labor.

The cranes that made cathedral building possible also weighed ship loads coming from early Cape Horn and Silk Road trade as the world became more global. Windmills were used as power sources for milling, sawing, and draining the Lowlands. All this was in place by the beginning of this time slice. Note that world exploration, including the New World, is also a fourteenth-century event.

In this context I want to focus upon what was clearly an instrumental revolution for the arts during the Renaissance. I will focus upon optics in the visual arts and the proliferation of musical instruments and innovative improvements for the musical arts. Prior to the Renaissance, the medieval arts had already acquired a greatly enlarged palette of colors: blues, greens, mauve, and especially gold. For music through the medieval period, which remained primarily vocal, instruments were largely restricted to secular street performances, where scores were produced. But within the fourteenth < seventeenth century slice here, virtually the entire range of instruments were invented, including: the sackbut (trombone), cornet, and trumpet, and other horns; winds, including recorders, flutes, and shawm (oboe); strings including the violin, viola, cello; organs; keyboards, including the harpsichord, etc.

What is usually taken as highly revolutionary in the Renaissance arts is the massive shift in visual style, the invention of ‘Renaissance Perspective’ and the ‘realism’ that goes with this style. My contention is that the optical technologies employed were primarily contributory to this shift. A primary device was the rediscovery of the camera obscura in the early Renaissance.³ The camera obscura had been known in China since –2,500 (it was described by Mo-Ti), and it was known to Arabs such as Al Hazen (who lived around 1038 AD or CE), who also was the author of the first optics theory.

Given the technological innovative thrust of the Renaissance, in its early days there was already a clear proliferation of optical devices. Leon Alberti used a large camera obscura to produce large paintings as early as 1437. Leonardo da Vinci used smaller cameras by 1508 and went on to describe an anatomy of the eye based upon this instrument. Other optical devices such as concave mirrors and the many Durer drawings of grid devices to aid artists are familiar to most of us. Perhaps the most famous use was that of Vermeer, as noted by Philip Steadman in *Vermeer’s Camera* (2001) and of late the documentary, *Tim’s Vermeer*, in which Tim Jenison attempts to produce a Vermeer painting using a camera obscura in a replicated Vermeer room.⁴ Simultaneously, the contemporary artist David Hockney published *Secret Knowledge* (2001), which revived a well-known history of camera obscura use in the Renaissance.⁵ What the camera—and to a lesser extent other drawing instruments—do is to provide an ‘automatic perspective machine’ by imaging a well-lighted three-dimensional scene, seen through or projected through an aperture or grid, and in the camera’s case produce a two-dimensional image (upside down) on a screen or white wall. Guiding grids in drawing can do the same thing, although less dramatically, and the artist produces the perspective image by drawing. By the late Renaissance more sophisticated optics such as the telescope and microscope appear, and these are used both by artists and early modern scientists such as Galileo. Note the parallel proliferation, from the vast instrumental inventions in music to the optical proliferation for the visual arts. As with the Paleolithic slice, each new instrument calls for new embodiment skills on the part of users. No one begins as a virtuoso. Today as I am myself taking my first watercolor classes, I realize that this absorptive medium is actually nearly the inverse of the

acrylic and oil media I previously used and which I had gained more mastery over than my still-primitive watercolor results.

Before leaving the Renaissance technological revolution, I want to briefly return to its musical instrument innovations. To produce resonance in volume, stringed instruments of great antiquity attached resonators. For example, Bushman cliff paintings, –4,500, show hunting bows using gourd resonators; these gourd resonators are also built into many Indian string instruments.⁶ Lutes, which go back at least to the tenth century AD or CE, turn out to produce only half the resonant volume of the violins of the sixteenth < seventeenth centuries. This is due to the “f” shaped holes of the late Renaissance, rather than the “O” shaped holes of lutes and earlier instruments.⁷

Slice 3: Contemporary Digital Technoart, Twentieth Through Twenty-First Centuries AD, CE

If we follow music history from the previous Renaissance time slice, much remains highly stable. Orchestral, chamber music, and much ‘classical’ music continue to use the same instruments—take Strativarius or Cremona violins as an example. Indeed, perhaps the biggest changes, mostly occurring in the early twentieth century, are shifts to *electric amplification* and the invention of *radio and recording devices*. Here a rapid development is clear, with rapid progression from Edison cylinders on to records, to vinyl, CDs, and today dominantly digital recordings, including MP3 and other digital formats.⁸ From early electric to contemporary digital electronics there is also a new revolution in sound production.

In the visual arts there is a correspondent technological revolution as well. ‘Modern art’ breaks all previous conventions. Already in the late nineteenth century the beginnings of impressionism and expressionism emerge, followed by Dada, surrealism, abstract expressionism and myriad other styles of visual presentation. But while these are radical stylistic innovations, painters are still using brushes and palette knives, but are also constructing with collage and sculpture-related productions. Major technological inventions though, transform visual practices. Photography, followed by cinema, which was early visual only, and later added music, then voice, to become the dominant audio-visual medium of the contemporary era. What this listing shows (which today is common knowledge) is that technologies change much more rapidly in the contemporary era. For example, I have already pointed out that Paleolithic art apparently did not use anything like a recognizable paint brush until late –12,000, and then it was crude. In previous work in the histories of writing technologies, brushes are associated with early calligraphy, possibly –5,000, but are also attributed to Chinese inventors as late as –2,800.⁹ As previously noted, the color palette remained very narrow, with some medieval additions and most expanded during the Renaissance (blues, greens, browns, and whites). Faster technological turnaround, exaggerated from the industrial revolution on, can again be illustrated by writing technologies. Quill pens dominated for nearly ten centuries, eighth to < early nineteenth AD or CE. The typewriter lasted roughly 120 years, although keyboards persist with electronic writing.

I shall conclude this slice with a brief look at musical synthesizers, not only because they are electronic, but because they are the most disjunctive compared to traditional acoustic instruments. The production of sound, the style of playing, the embodiment skills, and the sounds themselves are different. The oldest synthesizer usually is considered to be the *theremin*, a curious and accidentally developed instrument invented by Leon Theremin in 1920. This device consists of an “L” shaped antenna that projects an electronic field. Originally intended to be a device that could detect gas densities, Theremin discovered that when he placed his hands in the electronic field, weird sounds were emitted. Through some experimentation, he found that

he could play an ethereal type of music. (One documentary shows him teaching Lenin to play a Russian folk song—Theremin was a loyal Communist all his life and also developed a spy use for his instrument since it could both send—music—and receive.) This instrument did have an abbreviated music history as a single instrument, in some cases with an orchestra, but it also became popular as a background music source for science fiction movies.

A pioneer in sound studies, Trevor Pinch, has done a social history (with his co-author, Frank Trocco) of the two mid-century synthesizers, the Moog and the Buchla, in *Analog Days* (2002). Both these electronic instruments produce their own distinctive sounds, although much early work went into reproducing the sounds of standard instruments. As the authors point out, both synthesizer inventors had different approaches regarding sounds and the control of sounds in the playing technologies. Don Buchla favored unique synthesizer sounds and utilized a control panel with lots of buttons, tunable rheostats, and the like; Robert Moog favored a full range of sounds replicating other instruments, plus unique sounds, but adapted a standard piano-like keyboard for playing. As *Analog Days* points out, by using an already familiar keyboard, the Moog could call upon preskilled embodiment practices, which, gave the device a massive sales advantage. Its offspring, the various keyboard synthesizers, quickly dominated the market.¹⁰

By the end of the twentieth century, and now into the twenty-first, synthesizers have become digital. Sounds are electronically produced, and most machines utilize both sound production and a visualized, fully colored visual display. This audio-visual display is now typical for most media synthesizers, and it changes the experience of the controller-composer. Not only are the bodily skills required to play synthesizers very different from previous instruments, but the visualization technologies give sound a more immediate ‘spatial’ dimension. By changing the shape of the sound wave as displayed, the sound changes—and sounds can be produced that lie beyond the perceptual horizons of ordinary experience.¹¹

Although I have throughout suggested that each new technology, each instrument, calls for new embodiment skills, there is also an interesting periodic progression that can be noted. Beginnings suggest that embodiment is highly direct: it is hands on, from stencil signatures to clay sculpture formation. But the embodied use of tools—and art instruments from brushes to burins to flutes—is virtually simultaneous. Even, perhaps especially, the simplest instruments (such as a hunting bow used as a single-string instrument) call for highly honed skills to produce the overtones associated with such simple instruments. With the arrival of electronic and digital instruments there is a computer game–like emergence of eye-hand coordination. Today’s Apple iPhones, iPads, and other devices touch screens are tactile—or better. visual-tactile. Perhaps Buchla comes into his own after all?

Notes

- 1 See chapter 6 “Renaissance Art in Western Europe,” 170–221.
- 2 In discussing tools and weapons, Cook points to the difficulty of making spears straight. Spear throwers go back to this period and it is widely believed that these, atlatls, preceded archery. However, images of bows go back at least –15,000 years as I have shown in my research on the multistable history of the bow-under-tension. See my *Experimental Phenomenology: Multistabilities 2nd Edition* (SUNY Press, 2012) see chapter 13, “Variations on the camera obscura.”
- 3 Contrary to the Hockney-Falco claims that *Secret Knowledge* discovered the uses of optical devices by Renaissance artists, virtually all their illustrations from Alberti through Leonardo and on were summarized in the “Camera Obscura” entry in the 1929 edition of the *Encyclopedia Britannica*. “Camera Obscura” Vol 4, *Encyclopedia Britannica*, 1929 658 ff.
- 4 *Tim’s Vermeer*, a documentary release at the 2013 Toronto Film Festival, features Tim Jenison, who built a replica of Vermeer’s room setting for multiple paintings. Jenison used a camera obscura and related optics to reproduce a very viable likeness to a Vermeer painting.

- 5 David Hockney's *Secret Knowledge* was launched with much publicity, spawning documentaries, conferences and the like and too much opposition within the art history world. Yet, to historians of technology the thesis was not at all new—and some artists of the Renaissance wrote and admitted they used optics, including Leonardo da Vinci.
- 6 Sitar, drums, and other Indian instruments use very large gourds as resonators. Most European resonators are of wooden construction.
- 7 Nicholas Makris, an acoustical engineer at MIT, performed a multiyear study under a grant investigating this evolution. It is reported in *MIT News*, February 10, 2015, p. 1. See also my *Acoustic Technics* (Lexington, 2015), chapter 8.
- 8 See my parallel account of musical sound reproduction in “Sounds Beyond Sound” in Truax, et al., *Companion to Sounding Art* (2017, Routledge).
- 9 Dating is a major problem, as noted, with improved dating techniques many of the caves have turned out to be far older than earlier thought to be. I have found there is a similar problem with Chinese dating. Well-established dates, such as Mo-Ti and the camera obscura are no doubt real, but there is a tendency to credit China with ‘first discoveries’ that do not always hold up. Visiting China three times with lecture tours, I found many interlocutors claiming such firsts—for example, the Xian warriors complex, dated –2,500, has crossbows and archery that some claimed were Chinese first inventions. But the shaman holding a bow in the Trois Frères cave in France, dated –15,000, would seem to take precedence!
- 10 Leon Theremin, inventor of the theremin, 1896–1993.
- 11 It is my impression that while in a focal sense, we cannot hear (with our focal organs, the ears) these additions, they do affect the quality of the sound since hearing is actually a whole-body phenomenon. Recent physiology has reinforced phenomenology's notion of whole body perception. For example, taste nerve endings have recently shown that these are distributed throughout much of the body with more, numerically, elsewhere than on the tongue.

References

- Cook, J. (2013). *Ice Age Art: Arrival of the Modern Mind*. London: The British Museum Press.
- Galison, P. (1987). *How Experiments End*. Chicago: University of Chicago Press.
- Galison, P. (1997). *Image and Logic*. Chicago: University of Chicago Press.
- Hockney, D. (2001). *Secret Knowledge*. Oxford: Oxford University Press.
- Ihde, D. (1971). *Hermeneutic Phenomenology*. Evanston: Northwestern University Press.
- Ihde, D. (2007). *Listening and Voice*, 2nd edition. Albany: SUNY Press.
- Ihde, D. (2012). *Experimental Phenomenology, 2nd Edition: Multistabilities*. Albany: SUNY Press.
- Ihde, D. (2015). *Acoustic Technics*. Lanham: Lexington Books.
- Pinch, R., and Trocco, F. (2002). *Analogue Days*. Cambridge, MA: Harvard University Press.
- Steadman, P. (2001). *Vermeer's Camera*. Oxford: Oxford University Press.
- White Jr., L. (1962). *Medieval Technology and Social Change*. Oxford: Oxford University Press.

The Philosophy of Art and Technology

Jessie Mann

This chapter explores the ontological and teleological overlap between art and technology. Theories of technology will be reviewed alongside ontological and teleological theories of art. This chapter addresses digital art and the ways in which the resulting melding of technology and art changes the meaning and purpose of each of these phenomena and the embodiment of both. In doing so, I provide an overview of ideas and commentary on these topics in an effort to familiarize the reader with some of the current and historical conceptualizations of these phenomena and their overlap. Here, art is explored as the *purified essence* of transforming thought or existence into matter or action, the *essence* of the will-to-be; technology as the *force*, the *action*, the *being-in-the-world-ness* of this will-to-be. In what follows, it should become clear how it is just such a definition can be asserted.

Cyril Stanley Smith's *Art, Technology and Science: Notes on their Historical Interaction*, provides an excellent historical overview of the topic at hand. There, Smith argues that:

Artists have had far more intimate and continuing association with technology than they have had with science. In turn, the attitudes, needs, and achievements of artists have provided a continuing stimulus to technological discovery and, via technology, have served to bring to a reluctant scientific attention many aspects of the complex structure and nature of matter.

(Smith 1970: 494)

Smith's extensive historical research makes a compelling argument that, where science may be art's saner twin, technology is art's progeny. Heidegger takes a slightly different perspective, or, at least, he uses more poetic terminology. He argues that technology is the essence of bringing forth and revealing, a *coming to presence*. This coming to presence, he fears, is of a sort that makes demands of nature, that pulls from the materials of existence. Heidegger asserts that in this sense they are both a bringing forth. Heidegger posits that in sharing a root function, or ontology, with technology (that of bringing forth), art is uniquely positioned as a means of reflection upon technology. In Heidegger's conception we reveal ourselves to ourselves through our technology and art; that is, we bring ourselves forth into the world through these means. But in a way, this very revealing is an extension; it is an extension of self, a means of extending oneself so that reflection, or self-regard, is possible. In a Heideggerian framework, technology is an extension of humanity, it is the making manifest, the bringing forth of ourselves.

Just as technology can be conceptualized as an extension of self that facilitates a revelation of the self (collective or individual) to the self (collective or individual), so too can art. Both technology and art are extensions of an individual's thought, the translation of the information of their, and our collective, consciousness, into matter. Heidegger's technology makes demands of nature, it is a bringing forth of making and doing, an extension of intention and will, whereas

art is a making and doing regarding the very essence of making and doing. It is a bringing forth of the essence of *that we bring forth*. In the words of Joseph Campbell, "Nature as we know it is at once without and within us. Art is the mirror at the interface" (Banta 1989: 132). It is not the manifestation of will; it is the manifestation of the essence of Will itself, of Becoming, of Bringing forth. It is the will to be, making itself visible to itself.

Art is not about function in the way that technology is about function: art is concerned with the fact that we strive to convert thought/information into matter at all. Other than this, though, what defines art? For a concept that has received as much attention as it has, art is notoriously difficult to pin down. On both ends of the historical poles, art remains difficult to define, from the historical question "What was the first work of art?" to the current question, "Is this Art?" Nevertheless, like technology, art is one of the defining features of our humanity, as argued by the art theorist Lévy:

Art exists at the confluence of three main currents of virtualization and hominization: language, technology, and ethics (or religion). . . . It fascinates us because it implements the most virtualizing of all activities. Art provides an external form, a public manifestation, for emotions, for sensations that are felt in the innermost recesses of subjectivity. Although intangible and fleeting, we feel that these emotions are the spice of life. By making them independent of a particular time and place, or by providing them with a collective force (for the living arts at least), art enables us to share a way of feeling, a subjective quality of experience.

(Lévy 1998: 99)

So here we have it that art makes independent of time and place the intangible. It is the transformation of a subjective quality of experience into matter. Moreover, art, in Heideggerian terms, brings forth; it is *creation ex nihilo*, a "bringing to being." This *creation ex nihilo* aspect to art is very eloquently captured by the art theorist Roland Barthes in a review of the painter Cy Twombly's work:

The instrument of painting is not an instrument. It is a fact. Twombly imposes his materials not as something which will serve some purpose but as an absolute substance, manifested in its glory (in the theological vocabulary, God's glory is the manifestation of his Being). The materials are what the Alchemists called *material prima*—what exists prior to the division of meaning: a tremendous paradox, since in the human order nothing comes to man that is not immediately accompanied by a meaning, the meaning which other men have given it, and so on back to infinity. The painter's demiurgic power is that he makes the materials exist as substance; even if meaning emerges from the canvas, pencil and color remain "things", stubborn substances whose persistence in "being-there" nothing (no subsequent meaning) can annul.

(Barthes 1985:178)

Art is a means of bringing into being. It makes substantial something previously insubstantial. It is the breathing of some abstract 'persistence' into the materials of the world.

What, then, is being made substantial? What is being extended into the realm of matter? In some ways, it is our *selves* that are manifest. Recent literature has described technology as a virus that art, the cultural immune system, must keep in check (Blais, Joline, and Ippolito 2006: 9). Along these lines, and in a reversal of Smith's equation, technology/media theorist Marshall McLuhan asserts "art is precise advance knowledge of how to cope with the psychic and social consequences of the next technology" (McLuhan 1994: 66). In this case, art *follows*

from technology as a means of psychically coping *with* technology. Art is born of technology as a means whereby the collective mind can advance in a manner that allows it to survive the “blows” dealt it by the “extension of our faculties” (ibid.), as manifested by the latest technology. Indeed, many theorists have defined technology in extensionalist terms. In an excellent overview of this approach to technology, Philip Brey explores the extensionalist theories of Ernst Kapp, David Rothenberg, and McLuhan (Brey 2000). Each of these theorists, as well as Brey himself, conceptualize technology as an extension of one sort or another. For Brey, on one hand, “all artifacts” extend “the means by which human intentions are realized.” Rothenberg shares a similar perspective where technology aids in the execution of human will. Knapp, on the other hand, saw technology as a projection of human organs. McLuhan variously references extended bodies, nervous systems, cognitive functions, and consciousness as well (McLuhan 1994: 57). Mark Rowlands and Andy Clark similarly argue that technologies can extend our cognition. They cite low-fi technologies, like walking canes and notebooks, as well as hi-fi technologies, like smartphones, as extensions of the cognitive process.

Bill Viola, in describing the work of fellow video artist Peter Campus, not only describes this phenomena of the extension of the artist’s self, but he seems to sum up much of the nature of art making itself:

It [Campus’s camera] variously takes on the position of the artist himself, his reflection, an outside observer, a mental self-image, a double, an unknown protagonist, the room, an eye, a hand, an animal, an insect’s visual system. However, like a mirror of many facets all converging inward, the works keep returning to Campus himself and ultimately become a portrait of the Self searching for the ground of Being, peeling back layer upon layer of reality in the process.

(Viola 2010)

As Viola asserts, art making represents the point of view of the artist as well as the great Other, the abstract observer. The artists themselves are visible, but so too is the concept of being seen at all. So that, in another sense, what is being made substantial is something larger than just the artists themselves. What is being made substantial is the essence of that transition from abstraction to matter and back again. Art “symbolizes our will to shape and change the world and also puts forward the particular aspirations of its time” (Ascott 2003: 102).

One way that technologies and art differ is the sense of endurance. We assign a divinity to art; we believe a work of art to be ‘magic’ in a way that we don’t think of a tool as ‘magic.’ A study from *Topics in Cognitive Science* found that categorizing an object as “art” vs. “a tool” altered people’s sense of the persistence of these objects through time. The authors argue that these judgments about the continuity of art are somehow linked to our sense that art objects are “extensions of their creators” (Newman, Bartels, and Smith 2014). So that while both technology and art function to bring forth into being, that being which art brings forth is considered a more enduring extended being. The sense of persistence articulated by Barthes, cited earlier, seems to be one of the ways that art distinguishes itself from technology. Perhaps, from its very first instantiation, art has referenced this enduring quality. It has been argued that some of the first documented art objects were those decorative beads and pendants that were used in early burial rituals (Lewis-Williams 2010: 14) and that these burials “reflect a belief in the afterlife” (Dickson 1990: 95). It could be argued, then, that from its very beginning, art was an expression of the recognition of the translation between information and matter—a recognition that, though the body was gone, there was something immaterial that might remain. In the words of an artist, “Through the image is sustained an awareness of the infinite: the eternal within the

finite, the spiritual within matter, the limitless given form” (Tarkovsky 1987: 37). Though art has served many different functions over time, from its funerary birth to its pre-Renaissance iconography to modernism’s art for art’s sake, its meaning and purpose never stray far from a reference to the immaterial made manifest.

While art may have a unique element of endurance, a whiff of death avoidance, technology shares with it this quality of the immaterial made manifest. Technology is not devoid of its own form of escapism. Art may, by nature, conjure the sublime and the ever-enduring, but technology too references death in its push past reality, in its more recently salient push toward virtualization. As the language around technology has shifted toward the language of stored information and digitization, the ways in which technology too seems like a striving past death becomes more apparent. As Katherine Hayles writes:

In Moravec’s text, and at many other sites in the culture, *the information/matter dichotomy maps onto the older and more traditional dichotomy of spirit/matter*. The underlying premise informing Moravec’s scenario is the belief that an immaterial essence, which alone comprises the individual’s true nature, can be extracted from its material instantiation and live free from the body. As this wording makes clear, the contemporary privileging of information is reinforced by religious yearnings and beliefs that have been around for a long time and that are resonate with meaning for many people.

There are, of course, also significant differences between a mindset that identifies human being with the soul and one that identifies it with information. Spirituality is usually associated with mental and physical discipline, whereas the imagined escape of the soul-as-information from the body depends only on having access to the appropriate high technology.

(Masten, Stallybrass, and Vickers 1997: 186)

Technology, like art, pushes the spirit into the realm of matter. Both of these endeavors give form to the limitless, but art is more direct about what exactly it is doing.

It is this quality, this bringing forth, which reveals, in Heidegger’s language, or this virtualization, in more post postmodern language, that ontologically and teleologically binds art and technology. Technologies originally may not have been easily conceptualized as a translation of information into matter—or at least such a conceptualization was not as immediate as it was with art—but after the digital revolution, increasingly, we come to identify technology with abstractions, code, and purified information. Technology is revealing itself to be less the tools we make, and more abstract and purified information itself. Technology, like art, is revealing itself to be an extension of the will-to-reveal at work in the material realm. Furthermore, it seems, there is a reversal to the flow discussed earlier, for, as McLuhan put it, “In this electric age we see ourselves being translated more and more into the form of information” (McLuhan 1994: 57). There is an exchange, in both art and technology, mediated through a physical engagement with matter, of the self, that goes both ways. Via art and technology, we both make information material and we are material being made into information. Paul Levinson, a philosopher of technology, addresses the bidirectional flow thusly:

The act of extension, then, is an act of transformation, in which aspects of the universe bordering on the infinite and the infinitesimal are transformed into human proportion. Or perhaps the transformation works the other way, perhaps it is we who . . . are extended into the dimensions of the universe.

(as quoted by Rothenberg 1993: 25–26)

Whichever direction of flow it is, or whether both, the key ontological and teleological overlap between art and technology is the transformation between matter and information, material and spirit. In the words of the artist Joseph Beuys:

But [as an artist] one is forced to translate thought into action and action into object. The physicist can think about the theory of atoms or about physical theory in general. But to advance his theories he has to build models, tangible systems. He too has to transfer his thought into action, and the action into an object.

(Beuys 1990: 92)

A key component to the overlap under discussion is the handling of matter. On that topic, Cyril Smith makes the following point: “for some reason the technical side of art has been downgraded as ‘mere’ technique. Yet the handling of matter will always be necessary to give reality to the artist’s all-important vision” (Smith 1970: 545). Likewise, technology is often downgraded as the mere technique of science, but the handling of materials is essential to the manifestation of vision, essential to facilitating, in Heidegger’s words, “where truth happens.” The materials of translation are significant. This strive toward virtuality is about the *translation* between information and matter, not simply the information itself; the matter of which we speak is not inert and unimportant.

Despite the fact that the infinite/spirit may appear, by cultural standards, to be the more weighty of the two sides of this equation, the materials themselves, and the embodiment that moderates our engagement with these materials, are equally deterministic of the revealing that comes of these endeavors.

Worthy of special note is the way that artistic media are starting to converge with high technology. Digital art is in many ways further blurring the boundaries between art and technology, and is certainly redefining the nature of art. Computer-based art has been described by computer artist Myron Krueger as “a unique melding of aesthetics and technology in which creation is dependent on collaboration among the artist, the computer, and the participant” (Lopes 2010: 68). Indeed, a defining feature of digital art seems to be the distribution of authorship (Lévy 1998: 366–367).

This interactivity changes one of the primary dynamics in art, that of the exchange from the artist to the viewer. More than, that it changes the dynamic between the viewer and the object: “Using digital technology seems to give potential . . . to surround the body with a responsive and evolving environment” (Candy and Edmonds 2002: 142). Or, as artist Eric Heller sees it, digital art “allows us to create new media, with new rules, more naturally suited to the new tool” (Blais and Ippolito 2006: 40). These new rules, this new freedom, changes not only the artist-to-audience dynamic, and the viewer-to-object dynamic, but also the means of transmission and inspiration. As one theorist argues, internet art brings art viewing to a new context and brings new subject matter into the artistic context (Tresp). Furthermore, by bringing art out of the museum and gallery, and into a global information network, art has the potential to reach a much larger audience and to change the scope and connectivity of the dialog around art.

The artist as well as audience is made more connected: no longer do we see the solitary painter in the studio; rather art making is increasingly defined by collaboration. As the digital artist Jack Ox states: “Why is collaboration such an important part of the new century? With increased use of more and more specialized, complicated technology, it is far less productive to be an artist delivering a monologue alone in the studio” (Candy and Edmonds 2002: 53). So the digitization of art has changed being an artist and viewer alike from solitary existences to dynamic and interactive ones—and the division between the two roles, artist and viewer, itself is dissolved through the interactivity of digital art.

In this context, most importantly, this new art explicitly references its transformational action on information directly. Indeed, for some theorists, “transformation is the distinguishing feature of work in digital media” (Simanowski 2011: 24). As technology allows art to cross boundaries and engage material and embodiment in new ways, the work of art can more explicitly reference its core nature as the essence of this transformation between ‘the real and the virtual.’ The artist ORLAN, known for the surgical transformation of her own body as performance art, “describes her projects as belonging to both nature *and* technology, the real *and* the virtual” (Zylinska 2002). The digitization of art allows the ever-present extensionalism of art making to expand and be even more explicit in its self-reference. Art is now declaring itself to be the space of translation between the real and the virtual.

Also worthy of note is the unique interaction between art, technology, and the body. Art has paid special attention to the body, from early figurines, to the icon, to the portrait, to the performed body. Many technology theorists have also noted the ways that technology alters the body and our relation to it. When art is proposed as a way to process, to psychically respond to and prepare for, technological changes, as we have seen earlier, work like ORLAN’s or Stelarc’s then can be seen as a way to create a dialogue about the way technology modifies our bodies and the way that our bodies engage with the world.

Furthermore, these very body-modifying technologies also change what art is possible. As the artist Josephen Anstey writes in *We Sing the Body Electric: Imagining the Body in Electronic Art*:

The possibility of electrical augmentation of the body and of having virtual bodies attached to our real bodies suggests the freedom to transgress the normal limits of the body; limits of time and space, of appearance and fixed gender, of a unitary self, of self and other.

(as quoted in Wilson 2002: 154)

Just as digital art has expanded the role of both artist and viewer, current technologies are altering the very self that art extends. The self that artists imbues their work with is a self that is being extended and redefined by technology; the body that art has so often commented on and is born through an interaction with, is a body that is being forever altered by technology. The ‘contemporary body’

is the temporary actualization of an enormous hybrid, social, and technobiological hyperbody. [It] . . . moves outside itself, . . . is transmitted by means of a satellite, launches a virtual arm high in the air, flows through medical or communications networks. It entwines itself with the public body and . . . then returns, transformed, to its quasi-private sphere, and continues thus, sometimes here, sometimes there, sometimes alone or with others. One day, it will detach itself completely from the hyperbody and vanish.

(Lévy 1998: 44)

What art can be made with the body is altered by modern technology; the body itself is altered, and the cultural role of art is thereby altered as a result of these ongoing alterations. The artist is not only expected to comment on the technologized body, but also the artist may be called on to direct this technologizing. As art curator Jeffrey Deitch’s writes in the exhibit catalog accompanying the show “Post Human”:

As the organic, naturally evolving model of human life is replaced by the artificial evolution into the Post Human, art is likely to assume a much more central role. Art may have to fuse with science as computerization and biotechnology create further “improvements” on

the human form. Many of the decisions that will accompany the applications of computerized virtual reality and of genetic engineering will be related to aesthetics. Technology will make it possible to remodel our bodies and supercharge our minds, but art will have to help provide the inspiration for what our bodies should look like and what our minds should be doing.

(as quoted by Wilson 2002: 156)

According to this futuristic prescription, art functions as the imaginative arm of science and technology, responsible for dreaming up the realities to be made manifest through science and technology. Art—which, at its prehistoric birth in burial practices, hinted at a self disembodied but enduring—is now being handed the responsibility of dreaming up what possibilities of an enduring self there are.

So far in this chapter I have discussed extensionalism in the philosophy of art as well as technology, but with the remaking of identity and body just discussed, perhaps it must be recognized that “Rather than talk about the extensions of man we are faced with a discourse that articulates the inherently prosthetic character of human identity” (Zylinska 2002). It is possible, in their merging evolution, that art and technology have both come to represent pure transformation. The sense of an extended self, an extended body, an extended consciousness, gives way to an awareness that arising from their very nature, these concepts—self, body, consciousness—are extension itself. In short, the ‘extended self’ is a redundancy. Perhaps, the bringing forth that is art, with its enduring mark of the spiritual, can allay some of our fears about the way technology is going to influence human life in the coming years. As Billy Kluver, founder of E.A.T. (Experiments in Art and Technology), argues

artists have shaped technology. . . . They have helped make technology more human. They automatically will because they’re artists. That’s by definition. If they do something it automatically comes out human. There’s no way you can come out and say that if art is the driving force in a situation then it will come out with destructive ideas. That’s not possible. But what happens, of course, is that the artist widens the vision of the engineer.

(Candy and Edmonds 2002: 9)

As artists increasingly provide “inspiration for what our bodies should look like and what our minds should be doing,” as our vision widens, perhaps we will be able to reflect on ourselves in new and meaningful ways (Dietch, as quoted in Wilson 2002: 157). Indeed, it seems that the *essence* of the making and doing/revealing that is art is starting to mix with the *action* of making and doing/revealing that is technology, in such a way as to suggest a much broader philosophy. This new perspective—wherein we see our self as *prosthetic*, wherein we can deal in relativism and pluralism—reveals us to ourselves as the pure becoming made manifest, as reality and virtuality, as both information/spirit *and* matter. Through art and technology we are bridging the gap between these two realms and revealing ourselves to ourselves in the process. What is becoming clear is that we are translators, making information material, and the material into information. We are, again, “the mirror at the interface” (Banta 1989: 132).

References

- Ascott, R., and Shanken, E. A. (2003). *Telematic Embrace: Visionary Theories of Art, Technology, and Consciousness*. Berkeley, CA: University of California Press.
- Banta, T. (1989). *The Inner Reaches of Outer Space—Campbell, J.* Mahwah, NJ: Lawrence Erlbaum.
- Barthes, R. (1985). *The Responsibility of Form*. Berkeley, CA: The University of California Press.

- Beuys, J. (Compiled by C. Kouni) (1990). *Joseph Beuys in America*. New York: Four Walls Eight Windows.
- Blais, J., and Ippolito, J. (2006). *At the Edge of Art*. London: Thames and Hudson.
- Brey, P. (2000). Technology as Extension of Human Faculties. In C. Mitcham (ed.) *Metaphysics, Epistemology, and Technology. Research in Philosophy and Technology*, vol. 19.. London: Elsevier/JAI Press.
- Candy, L., and Edmonds, E. (2002). *Explorations in Art and Technology*. Great Britain: Springer.
- Dickson, D. B. (1990). *The Dawn of Belief: Religion in the Upper Paleolithic of Southwestern Europe*. Tucson, AZ: University of Arizona Press.
- Lévy, P. (1998). *Becoming Virtual: Reality in the Digital Age*. New York: Plenum Press.
- Lewis-Williams, J. D. (2010). *Conceiving God: The Cognitive Origin and Evolution of Religion*. London: Thames & Hudson.
- Lopes, D. M. (2010). *The Philosophy of Computer Art*. New York: Routledge.
- Masten, J., Stallybrass, P., and Vickers, N. (1997). *Language Machines: Technologies of Literary and Cultural Production*. New York: Routledge.
- McLuhan, M. (1994). *Understanding Media: The Extension of Man*. Cambridge, MA: The MIT Press.
- Newman, G. E., Bartels, D. M., and Smith, R. K. (2014). Are Artworks More Like People Than Artifacts? Individual Concepts and Their Extensions. *Topics in Cognitive Science*, 6(4), 647–662.
- Rothenberg, D. (1993). *Hand's End: Technology and the Limits of Nature*. Berkeley, CA: University of California Press.
- Simanowski, R. (2011). *Digital Art and Meaning: Reading Kinetic Poetry, Text Machines, Mapping Art, and Interactive Installations*. Minneapolis, MN: University of Minnesota Press.
- Smith, C. S. (1970). Art, Technology, and Science: Notes on Their Historical Interaction. *Technology and Culture*, 11(4), 493–549.
- Tarkovskii, A. A. (1987; 1986). *Sculpting in Time: Reflections on the Cinema*, 1st American edition. New York: Alfred A. Knopf.
- Tresp, L. *Digital Art: Chicago School of Media Theory*. Available at: <https://lucian.uchicago.edu/blogs/mediatheory/keywords/digital-art/>
- Viola, B. (2010). *Peter Campus: Image and Self*. New York: Brant Publishing.
- Wilson, S. (2002). *Information Arts: Intersections of Art, Science, and Technology*. Cambridge, MA: MIT Press.
- Zylinska, J. (ed.) (2002). *The Cyborg Experiments: The Extension of the Body in the Media Age*. New York: Continuum.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Political Space and Agency



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Robots With Guns

Shannon Vallor

Introduction

The development and use of armed robots for military and law enforcement purposes is one of the most ethically contentious and politically challenging issues of the early twenty-first century, one that increasingly commands attention not only from philosophers of technology but also from citizens, media and watchdog groups, thought leaders, and institutions of law, governance, and public policy from around the globe. Responding to mounting international concern, a 2013 report commissioned by the United Nations recommended “that States establish national moratoria on aspects of LARs [lethal autonomous robots]” (Heyns 2013). In 2015 the Future of Life Institute released an open letter signed by tens of thousands of scientific researchers and other stakeholders urging a global “ban on offensive autonomous weapons beyond meaningful human control” (Russell et al. 2015). Yet the multi-billion dollar global market in defense robotics is projected to grow rapidly in coming decades as nations and private law enforcement agencies bet on the promise of armed, and increasingly intelligent, robots.

The philosophical debate over armed robots is grounded in several overlapping areas of research, including machine ethics and roboethics, military ethics (especially Just War Theory) and the philosophy of cognitive science and artificial intelligence (AI). Philosophical questions driving this debate include, but are not limited to, the following:

1. Could it ever be morally acceptable for a machine to autonomously kill a human being, or for a person to design or use a machine with this capability? Or are such weapons *mala en se* (intrinsically wrong), as with rape and biological weaponry?
2. When we speak of autonomous lethal robots, what qualifies as ‘autonomous’? Should humans be kept ‘in the loop,’ or ‘on the loop’? What qualifies as ‘meaningful human control’ over the actions of armed robots?
3. Can the use of armed robots be justified by consequentialist arguments, for example, if they reduce risk to noncombatants, shorten the duration of armed conflicts, or reduce the number of unlawful war atrocities committed by humans?
4. Is it ethical or prudent to make war less costly? Might this lower the psychological, political, and economic barriers to starting or continuing wars?
5. Could armed robots ever display the powers of discrimination and proportional judgment mandated by international laws of armed conflict (LOAC), international humanitarian and human rights law (IHL/IHRL), and Just War Theory?
6. Could artificially intelligent armed robots ever be held morally or legally responsible for their conduct in war? Does their use violate due process?
7. Can ethical concerns about armed robots be alleviated by developing robots with nonlethal offensive capacities? Or by restricting armed robots to defensive use?
8. Could armed robots pose an ‘existential threat’ to human existence or to the planet, perhaps by acquiring intelligence or other strengths superior to those of their creators?

Armed Robots and Autonomy

The development of armed robots is not entirely new. If we adopt a very broad construal of the term *robot* that requires only the ability to sense, calculate, and respond physically in its local environment (Lin et al. 2008: 4), then armed robots have existed for decades in the form of defensive technologies such as the U.S. Phalanx antimissile system and Israel's 'Iron Dome.' Another example of an armed robot already in use is South Korea's SGR-1 sentry system, which can identify and fire upon human targets detected within the demilitarized zone. Because defensive systems often require a response to incoming threats on very short timescales, such systems are also designed to be able to operate with some degree of independence from human operators: that is, they can identify and fire upon an incoming target without prior approval from a human operator, if that mode of operation is desired.

A related group of technologies are unmanned aerial, land, and underwater vehicles or 'drones,' which may be armed or unarmed. Armed drone systems remotely controlled and fired by human pilots, such as the U.S. Predator UAV (unmanned aerial vehicle), do not, under most conventional definitions, qualify as robots. However, UAVs such as the United Kingdom's Taranis and the United States' X47-B are capable of increasingly automated flight operations, and some drone systems are capable of automated armed response, as in the case of Israel's Harpy/Harop 'fire and forget' antiradar UAV. The consensus among robotics and military experts is that current trends predict increasing automation in armed robotic systems, including the capacity for autonomous targeting of and application of lethal force to human beings (Singer 2009).¹ It is this latter development that generates the greatest public controversy, and the most challenging philosophical questions.

The most pressing moral question is whether it can ever be ethical to design or build a robot that can kill a human on its own power and authority. Some scholars assert that lethal autonomous robotic systems (LARS) qualify as *mala en se* uses of force, as with rape, genocide, and biological and chemical weapons (Wallach 2015). The supposed 'evil' in this case arises from the distinctly inhuman agency of the machine allowed to impose lethal force. Since robots currently do not (and may never) possess a genuinely moral understanding of the value or dignity of human life, or the capacity for moral emotions such as empathy, compassion or guilt, or a robust sense of personal responsibility to others, it is often argued that allowing them to make independent judgments about lethal force against humans would violate certain fundamental rights of their victims, for example, the right to be considered as moral ends in themselves (Asaro 2012). Those who hold this view suggest that LARS may violate the *Martens Clause* of international law, which states that novel means of armed conflict must be compatible with "the principles of humanity and the dictates of the public conscience" (Meron 2000).

Among those who favor the development of autonomous lethal robots, one response is to assert that such systems can eventually be programmed with sufficient ethical awareness and self-governance to qualify as artificial moral agents (AMAs) capable of the responsible use of lethal force, performing at least as ethically—and likely more ethically—under battle conditions than do human soldiers (Hall 2007; Lin et al. 2008; Arkin 2009, 2010). The design of AMAs is a central topic in machine ethics and roboethics that extends well beyond military applications. Most researchers in this field expect AMAs for the foreseeable future to have only *functional* morality, that is, reliably successful autonomous moral behavior in specific practical contexts (Moor 2006; Wallach and Allen 2009; Sullins 2010). This is a higher standard than that of mere *operational* morality, which involves preprogrammed behavioral scripts the machine is required to follow. A *functional* moral agent would be able to exercise decision-making power of its own, using its own assessment of the moral situation (Lin et al. 2008: 26). Yet this would

still be short of full moral agency of the sort possessed by humans, with our comprehensive understanding and embodiment of human intentions, emotions, values, and social norms.

Still, the prospects of successfully engineering even a functional moral agent are uncertain, and thus the ongoing debate over AMAs brings a contemporary urgency to long-standing philosophical questions about the nature of moral agency, responsibility, accountability, and autonomy (Wallach and Allen 2012). Finally, there is the glaring problem of which ethical frameworks or theories are appropriate to use as the normative standard for robotic behavior, given the wide range of cultural and philosophical perspectives on morality. Yet the debate over AMAs cannot be dismissed as merely academic or speculative, for there are immediate practical stakes of great public import; already, defense agencies have committed tens of millions of dollars to research and development of artificial moral governance systems (Tucker 2014).

One of the basic assumptions of advocates of armed robots is that they will be more trustworthy than armed humans, who have an extremely poor track record in the ethical uses of force, even in highly socialized domains with well-developed codes of conduct (Arkin 2009; Sullins 2010). Even ‘normal’ humans are subject to deeply embedded—often unconscious and irrational—thought patterns driven by fear, disgust, anger, greed, lust, ignorance, laziness, sectarian resentment, and bias. Add to that the significant proportion of humans subject to sociopathic or other pathological tendencies toward violence, and robots with guns start to look like a fairly attractive substitute for people with guns.

Yet how reasonable is it to think that armed robots will be free of our dangerous human biases, fears, and blind spots, since we will be the originators of their code? In order to give robots the power to understand the human situations in which force is to be appropriately used, we must lend them *our* understandings of what counts as ‘good’ or ‘appropriate’ behavior, of who is ‘suspicious’ and who is ‘trustworthy,’ who appears ‘aggressive’ or ‘benign,’ who counts as a ‘perpetrator’ and who as a ‘victim.’ In addition to the fact that we are notoriously prone to having these distinctions infected by our own cognitive and moral shortcomings, it is important to ask *who* will be the ‘we’ I speak of, the ‘we’ that will craft and test the moral sensibilities of armed machines. The technologies that make armed robots possible are *already* concentrated in the hands of a very few, and robotics research and development is funded by a relatively small number of private corporations, research universities, and military agencies whose leaders and policy makers are overwhelmingly male, White, wealthy, and powerful, often having vested interests that do not necessarily correspond to those of humanity as such.

One response to the moral dilemmas and risks associated with armed robotic systems is to propose that they remain restricted to semi-autonomous modes of operation, that is, operations that retain a human presence ‘in the loop’ of command and control. A weakened version of this is the human ‘on the loop,’ in which the robotic system can execute its functions entirely without human assistance or input, yet humans are retained in a supervisory role and can intervene if deemed necessary.

One worry about this solution is that humans are known to overattribute mental states and capabilities to robotic systems on the basis of relatively thin behavioral similarities with intelligent humans (Scheutz 2012). If great care is not taken, human supervision of LARS may well devolve to ‘rubber-stamp’ approvals and post-hoc justifications of robotic actions initially presumed to be wise. This worry is magnified by the fact that expensive, high-tech systems beg to be used: what good is a robotic agent whose judgments you can’t trust? To a colonel with a fancy robot plane that ‘thinks’ faster and ‘sees’ farther, will anyone in the robot’s crosshairs be accepted as a target to be destroyed? Finally, it remains unclear for how long *meaningful* human oversight will remain feasible in the light of ongoing advancements in the speed, precision, and scale of armed robotic systems. As the speed and powers of such systems surpass those of

their human monitors, ‘loop’ protocols will be increasingly difficult to maintain without loss of system effectiveness and competitive advantage (Sharkey 2012, 2008; Vallor 2013).

The Case for Armed Robots

While important philosophical objections to the design and use of armed robots have been voiced in scholarly and public fora, there are equally important philosophical arguments offered in favor of the use of armed robots. The theoretical basis of such arguments is typically *utilitarian*: If armed conflict cannot always be avoided, and if the development and use of armed robots can save more human lives and/or prevent more suffering than other methods of fighting, then, it is argued, we have a moral duty to develop and use such technologies. In particular, armed robots are generally invoked as a means of eliminating the kinds of moral atrocities and errors produced by human biases and emotions in armed conflict—atrocities and errors that have remained virtually impossible to eliminate by any other practical means (Arkin 2009, 2010). Advocates for armed robots argue that by minimizing the role of human emotion and irrationality in armed conflict, we can expect the human cost and even the duration of such conflicts to be effectively minimized.

Objections to this line of argument vary. One can accept the basic utilitarian conditional but deny that the consequences of armed robots are likely to be as salutary as their advocates predict. For example, it is commonly pointed out that the high cost and ugliness of armed conflict are the strongest deterrents to its practice (Sharkey 2012). If making armed conflicts less individually terrible and costly makes armed conflict *in general* a less-forbidding prospect, then it may be chosen more frequently as a means of resolving human disputes (Sparrow 2007). In this scenario, the use of armed robots may in fact *increase* overall human suffering in the long run. One may also question whether armed robots are likely to be as reliably beneficial to their makers as their advocates expect, given the inevitability of unanticipated software bugs and failures, and the vulnerability of machine software to exploitation by hackers and terrorists. A different objection is to hold the utilitarian calculus as morally insufficient, by asserting that fundamental human rights or moral duties are violated by the use of armed robots (Asaro 2008, 2012), or that the widespread use of armed robots is likely to undermine human character and flourishing in unacceptable ways (Sparrow 2013; Vallor 2014).

Armed Robots and Just War Theory

Many concerns about the compatibility of armed robots with IHL/IHRL and LOAC are conceptually rooted in the philosophical history of Just War Theory. The dictates of Just War Theory of particular concern in the case of armed robots are the principles of *distinction*, *proportionality*, and *necessity* (Asaro 2008; Sharkey 2012).

The principle of distinction, embedded in the Geneva Conventions, finds its roots in the philosophical tradition that holds that civilians or noncombatants must not be targeted or indiscriminately harmed in the conduct of a just war. Thus the cognitive capacity of *discrimination* is closely connected with the principle of distinction. Advocates of armed robots have claimed that sufficiently sophisticated programming will allow robots to be equipped with this capacity. They assert that since humans discriminate between combatants and noncombatants using a combination of available perceptual data and prior knowledge, robots with sophisticated perceptual and memory systems should soon be able to accomplish the same with a performance comparable or superior to that of human soldiers (Arkin 2009). By programming in rules that dictate avoidance of harm to noncombatants when possible, and giving robots sufficient training in discriminative exercises, the moral obstacle presented by the principle of distinction may be reduced to a tractable engineering problem.

Critics point out that the conceptual distinction between combatant and noncombatant remains a highly contextual, cultural, and situationally variable one that we humans navigate using our native ability to read the mental states of others, and thus may well be irreducible to a rule-based algorithmic decision procedure (Guarini and Bello 2012; Roberts 2011). However, this point may come under pressure from recent advances in machine ‘deep learning’ techniques that show promise in training computers to perform discriminative and classification tasks, even without being supplied rules or fixed criteria in advance.

The requirements of the principles of proportionality and necessity yield a similar debate. The principle of proportionality holds that the use of force in armed conflict must be appropriately modulated; the scale, method, and timing of the use of force should be carefully selected in order not to exceed what is required by reason, and in a manner sensitive to the value of civilian life and property. The use of force must also itself be *necessary*: it must accomplish a legitimate military objective and not be an act of violence for its own sake. For example, if my unit’s mission can be accomplished just as well and as safely by a narrowly targeted raid on a munitions bunker or a command and control station, it would be unjust and unlawful for me to order a massive airstrike on the entire port, especially if the military value of the port is low relative to its civilian value. Again, advocates of armed robots will claim that these highly complex and contextual judgments of military necessity and proportionality are reducible to challenging but tractable tasks of software engineering, while their detractors insist that such judgments require a kind of moral understanding and practical wisdom inaccessible to nonhuman systems (Sharkey 2009). Others have argued that even *if* machines could someday perform these tasks as well or better than humans, we have independent reasons to want humans to continue to perform them (Asaro 2012; Vallor 2013).

Additionally, armed robots raise important questions about responsibility, transparency, and due process. Who will be held responsible when an armed robot fails, as all technological systems inevitably do? How will a civilian injured or bereaved by an errant robot find justice? This is especially problematic given that advanced machine learning techniques do not readily yield to human inspection, nor can they be straightforwardly converted to an account of ‘reasons’ why a certain result was reached or an action taken. The layers of deep learning systems are largely opaque to human understanding, yet the right to know the *reasons* behind a legal decision is an essential part of due process protections. Since robotic systems that can be morally accountable for themselves and their reasons are nowhere near being developed, and since robots cannot be meaningfully punished for an unjust decision, their use in the application of force without direct and meaningful human control is held by many scholars to be deeply problematic (Sparrow 2007; Sharkey 2010). Others resist this conclusion, asserting that robots *could* be constructed with the affective capacity to suffer punishment if that were deemed desirable, and furthermore, that a necessary link between punishment and justice is questionable at best (Lokhorst and van den Hoven 2012). Alternatively, if one holds that justice is necessarily and in essence a moral bond or contract between humans, then it is not an empirical question at all whether robots can be just—a human can receive justice *only* from another human. To use machine delegates as substitutes for our place in a judicial process would be to “dehumanize justice, and ought to be rejected in principle” (Asaro 2012: 701).

Nonlethal and Defensive Use of Armed Robots

Not all uses of armed robots are equally controversial. For example, some scholars strongly opposed to lethal autonomous robotics systems are willing to accept the use of automated systems that target nonhumans such as missiles and radar installations, or lethal semi-autonomous robots that remain under meaningful human control. Some hold that defensive uses of armed robots can be morally distinguished from offensive ones, while others regard the distinction

between defensive and offensive action as too fluid to be a reliable safeguard. A defensive armed robot is on the defense only when it hasn't made a mistake; to the person it kills or maims in error, or whose behavior toward the robot was ambiguous in its aggression, the machine has applied offensive force.

It is also uncertain whether expanding worldwide investment in armed robotic systems will make the restrained use of such systems increasingly difficult. Much of the debate about armed robots thus reduces to different estimations of the likelihood of one or more 'slippery slopes' in their use. For example, how likely is it that armed robots will remain out of the hands of non-state actors or rogue nations who regard themselves as not bound by international law or just war principles? How plausible is it that commitments to strictly defensive or nonlethal uses, or strong 'on the loop' protocols, will hold fast in the face of growing military threats from nations that eschew such restraints? Finally, how likely is it that armed robots equipped with increasingly sophisticated forms of AI will remain obedient, predictable, and beneficial to their makers? Here Asimov's Three Laws of Robotics, and their well-known fictional insufficiencies, serve as a prescient illumination of the challenges of machine ethics (Anderson 2008).

The uncertainty embedded in these scenarios points to the importance of a carefully considered theory of ethical *risk assessment* in the creation of policy concerning the development and use of armed robots (Lin et al. 2008). How should the risks be calculated? What kinds of harms should be considered, and to whom, and how should these be weighted? Which global stakeholders should be consulted in determining acceptable levels of risk, and what about those stakeholders who are politically or physically prevented from participating in the assessment? What qualifies as informed consent in open public discussion of the risks, especially given the obstacles to full transparency posed by national security, military secrecy, and intellectual property interests? How should unquantifiable but grave risks posed by armed robots to future generations be balanced against more determinate but limited benefits of their use in the present?

Armed Robots and Existential Risk

The importance of appropriate risk assessment is further underscored by the conclusion of many that the development of armed robots poses some *existential risk* to humanity, that is, a risk of endangering meaningful human existence on this planet. There are multiple versions of this claim. For example, armed robots might pose an existential risk by destabilizing the traditional balance of powers or modes of deterrence to war, increasing the chances of a worldwide armed conflict that wipes out human civilization or existence. Perhaps the most dramatic and hotly debated possibility—though by no means the most likely one—is that the union of robotics and AI will create a form of intelligent life so far above and unlike our own that human values and existence will be viewed by such beings as unworthy of ongoing protection, regardless of their original programming. This 'Terminator' or 'Skynet'-scenario, as it is sometimes colloquially referred to, is often linked to theories about the Singularity—a hypothesized point in history at which incremental improvements made by humans to artificially intelligent systems begin to feed back upon themselves and boost machine intelligence exponentially, taking the development of AI out of human hands and creating a 'superintelligence' beyond our comprehension or control (Chalmers 2010; Bostrom 2014). Imagining a superintelligence equipped with lethal weaponry clearly magnifies the existential worry.

Fears of superintelligence have received a great deal of popular media attention of late, partly due to their articulation by well-known and highly respected public figures such as inventor Elon Musk and cosmologist Stephen Hawking, and partly due to their echoes of early twenty-first century apocalyptic themes in popular entertainment. Such worries motivate many calls to ban the development of autonomous lethal robots, such as the Future of Life open letter backed

by Musk and Hawking. They also inspire calls to invest more quickly and heavily in the development of ‘Friendly AI’ with human-safe values ‘baked in’ (Yudkowsky 2008).

Yet many philosophers of technology focus on more sober existential worries about armed robots than the ‘Singularity.’ Instead of the worry that robots will become too smart for us, such scholars conclude that the far greater risk is that smart humans will rely too heavily on robots that aren’t smart *enough*, robots whose capacities for intelligent decision making we overestimate at our peril (Veruggio and Abney 2012; Wallach 2015). While the superintelligence hypothesis is wildly speculative, we have clear empirical evidence that automated agents can malfunction in unpredictable ways, potentially wreaking large-scale havoc on human institutions, such as the Flash Crash of the stock market in 2010 and the Knight Capital Group stock meltdown in 2012, both involving glitches in the automated behavior of high-frequency trading algorithms. A ‘flash crash’ that affects an entire army of autonomous robot soldiers, or a large squadron of heavily armed and fully autonomous drone bombers, is a nightmare scenario that prudence demands we find ways to prevent. Indeed, dangerous malfunctions of armed robotic systems have already occurred on several occasions, including the 2007 case of a semi-autonomous cannon in South Africa that killed nine and wounded fourteen friendly observers (Lin et al. 2008: 7). Related scenarios involve the hacking or theft of insecure robotic systems by malicious human actors. The question is whether such scenarios can be adequately prevented without restricting or banning the development of these technologies in the first place.

Yet another more sober and empirically grounded fear than that of superintelligent robots is the potential use of armed robots to consolidate political control in the hands of a very few international powers or individuals, increasing the risk of local and even global tyranny. The potential for use of armed robots in law enforcement, crowd control, and private security and surveillance, for example, has not been given the same share of consideration as has their use in international conflicts. How will citizens respond to the presence of armed robots patrolling public spaces? How will that change how publics behave, or how they perceive and think about political and legal authority? Are armed robots a viable solution to the increasing challenge of policing in contexts infused with racial bias and injustice? Or would this be the next step toward dehumanizing citizens as objects to be controlled from a safe distance, rather than persons to be engaged and understood? The risk that despots will use such technologies to more effectively suppress dissent or popular resistance is obvious and grave—especially if these technologies will be inaccessible to ordinary citizens without considerable wealth, and inoperable by those without advanced education in software and mechanical engineering. For this reason, understanding the ethical *and* political valence of armed robotic systems will be an increasingly important research area for twenty-first century philosophers of technology.

Note

- 1 It is common to distinguish between *automated* and *autonomous* robotic operations; the latter involves independent decision-making power, in this context, with respect to lethal use of force (Asaro 2012).

References

- Anderson, S. L. (2008). Asimov’s ‘Three Laws of Robotics’ and Machine Metaethics. *AI and Society*, 22(4), 477–493.
- Arkin, R. (2009). *Governing Lethal Behavior in Autonomous Robots*. Boca Raton: CRC Press.
- Arkin, R. (2010). The Case for Ethical Autonomy in Unmanned Systems. *Journal of Military Ethics*, 9(4), 332–341.
- Asaro, P. (2008). How Just Could a Robot War Be? In P. A. E. Brey, A. Briggle, and K. Waelbers (eds.) *Current Issues in Computing and Philosophy*. Amsterdam: IOS Press, 50–64.

- Asaro, P. (2012). On Banning Autonomous Weapons Systems: Human Rights, Automation and the Dehumanization of Lethal Decision-Making. *International Review of the Red Cross*, 94(886), 687–709.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Chalmers, D. (2010). The Singularity: A Philosophical Analysis. *Journal of Consciousness Studies*, 17, 7–65.
- Guarini, M., and Bello, P. (2012). Robotic Warfare: Some Challenges in Moving From Noncivilian to Civilian Theaters. In P. Lin, K. Abney, and G. A. Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, 129–144.
- Hall, J. S. (2007). *Beyond AI: Creating the Conscience of the Machine*. New York: Prometheus Books.
- Heyns, C. (2013). Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. *United Nations General Assembly, Human Rights Council*. Pub. A/HRC/29/37, April 9.
- Lin, P., Bekey, G., and Abney, K. (2008). *Autonomous Military Robotics: Risk, Ethics and Design*, Report for U.S. Office of Naval Research. San Luis Obispo, CA: California Polytechnic State University. Available at: http://ethics.calpoly.edu/onr_report.pdf.
- Lokhorst, G., and van den Hoven, J. (2012). Responsibility for Military Robots. In P. Lin, K. Abney, and G. A. Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, 145–156.
- Meron, T. (2000). The Martens Clause, Principles of Humanity, and Dictates of Public Conscience. *The American Journal of International Law*, 94(1), 78–89.
- Moor, J. (2006). The Nature, Importance and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21(4), 18–21.
- Roberts, A. (2011). The Civilian in Modern War. In H. Strachan and S. Scheipers (eds.) *The Changing Character of War*. Oxford: Oxford University Press, 357–380.
- Russell, S., et al. (2015). Autonomous Weapons: An Open Letter From AI & Robotics Researchers. Future of Life Institute. Available at: http://futureoflife.org/AI/open_letter_autonomous_weapons#signatories.
- Scheutz, M. (2012). The Inherent Dangers of Unidirectional Emotional Bonds Between Humans and Social Robots. In P. Lin, K. Abney, and G. A. Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, 205–222.
- Sharkey, N. (2008). Cassandra or the False Prophet of Doom: AI Robots and War. *IEEE Intelligent Systems*, 23(4), 14–17.
- Sharkey, N. (2009). Death Strikes From the Sky: The Calculus of Proportionality. *IEEE Science and Society*, 28, 16–19.
- Sharkey, N. (2010). Saying ‘No!’ to Lethal Autonomous Targeting. *Journal of Military Ethics*, 9(4), 369–383.
- Sharkey, N. (2012). Killing Made Easy: From Joysticks to Politics. In P. Lin, K. Abney, and G. A. Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, 111–128.
- Singer, P. W. (2009). *Wired for War: The Robotics Revolution and 21st Century Conflict*. New York: Penguin Group.
- Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, 24(1), 62–77.
- Sparrow, R. (2013). War Without Virtue? In B. J. Strawser (ed.) *Killing By Remote Control: The Ethics of an Unmanned Military*. New York: Oxford University Press, 84–105.
- Sullins, J. (2010). RoboWarfare: Can Robots Be More Ethical Than Humans on the Battlefield? *Ethics and Information Technology*, 12(3), 263–275.
- Tucker, P. (2014). Now The Military Is Going to Build Robots That Have Morals. *Defense One*, May 13. Available at: www.defenseone.com/technology/2014/05/now-military-going-build-robots-have-morals/84325/.
- Vallor, S. (2013). The Future of Military Virtue: Autonomous Systems and Moral Deskillling in the Military Profession. In K. Podens, J. Stinissen, and M. Maybaum (eds.) *Proceedings of 5th International Conference on Cyber Conflict (CyCon 2013)*. Tallinn, Estonia: NATO CCDCOE, 471–486.
- Vallor, S. (2014). Armed Robots and Military Virtue. In L. Floridi and M. Taddeo (eds.) *The Ethics of Information Warfare*. Switzerland: Springer, 169–185.
- Veruggio, G., and Abney, K. (2012). Roboethics: The Applied Ethics for a New Science. In P. Lin, K. Abney, and G. A. Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, 347–363.

- Wallach, W. (2015). *A Dangerous Master: How to Keep Technology From Slipping Beyond Our Control*. New York: Basic Books.
- Wallach, W., and Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.
- Wallach, W., and Allen, C. (2012). Moral Machines: Contradiction in Terms or Abdication of Human Responsibility? In P. Lin, K. Abney, and G. A. Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, 55–68.
- Yudkowsky, E. (2008). Artificial Intelligence as a Positive and Negative Factor in Global Risk. In N. Bostrom and M. M. Čirković (eds.) *Global Catastrophic Risks*. New York: Oxford University Press, 308–345.

Educational Technology

Andrew Wells Garnar

Alongside the hype that surrounds the accelerated introduction of and the increased use of information and communication technologies (ICTs) are fears about what these technologies do to education. Norm Friesen's "Brilliance and Simulation" offers a particularly vivid account of these concerns through a phenomenological exploration of the different ways to dissect frogs. On the one hand, there is the traditional way that involves the cutting open of a formaldehyde-preserved frog. On the other, there are computer-based virtual frog dissection programs. While Friesen is careful to limit his analysis to a comparison of the different experiences produced by these two sorts of dissection practices, rather than wading into thorny questions about the ethics of virtual and real frog dissections, he does end with a hint of criticism of the virtual. While the appeal of virtual dissection is that students do not have to work with a dead animal (which is a definite draw for those with objections to the practice), this is also a drawback. The simulation cannot fully replicate, or arguably approximate, the smell of the chemicals, the crack of the rib cage, the difficulty of removing the organs, etc. To describe this simulated nature, Friesen introduces Albert Borgmann's analysis of the "hyperreal" from *Crossing the Postmodern Divide*. There, Borgmann argues that 'postmodern' technologies, in particular computers, present a reality that is more brilliant and flawless than 'real' reality, what also might be called *world* in something like a Heideggerian sense. Through the presenting a simulation of the world, the hyperreal highlights certain features far more than they would be experienced otherwise. In doing so, the simulation is more perfect than the real thing because it presents it cleanly, without friction. Without work on the teacher's part, the simulation lacks important aspects that important for education, specifically: "Opacity, disruption and upheaval—rather than . . . withdrawal and protection" (Friesen 2011: 198). The former occurs in 'real' dissections, while virtual dissections make the latter far more likely. The tendency of virtual dissection to limit opacity, disruption, and upheaval prevents it from being more than a mere simulation, even if such simulations can still be used to achieve certain pedagogical goals. While this is a narrow example of how ICTs transform education, Friesen, building in part of Borgmann, effectively captures the risk of using simulations.

Yet, concern about what technology will do to education is nothing new. Allan Collins and Richard Halverson usefully catalogue some of the criticisms American educators have offered over the centuries of cutting-edge technologies:

From a principal's publication in 1815: "Students today depend on paper too much. They don't know how to write on a slate without getting chalk dust all over themselves. They can't clean a slate properly. What will they do when they run out of paper?"

From the journal of the National Association of Teachers, 1907: "Students today depend too much upon ink. They don't know how to use a pen knife to sharpen a pencil. Pen and ink will never replace the pencil."

From *PTA Gazette*, 1941: “Students today depend on these expensive fountain pens. They can no longer write with a straight pen and nib. We parents must not allow them to wallow in such luxury to the detriment of learning how to cope in the real business world which is not so extravagant.”

From *Federal Teachers*, 1950: “Ballpoint pens will be the ruin of education in our country. Students use these devices and then throw them away. The American values of thrift and frugality are being discarded. Businesses and banks will never allow such expensive luxuries.”

(Collins and Halverson 2009: 30–31)

Reviewing criticisms like there is useful because it helps to shake up some of the contemporary debates about educational technology. Most of the present debates focus on the role of computers in education. While not denying the importance of such debates, it is helpful to remember that *educational technology* does not only refer to computers, or even electronic devices generally. Instead, the term can potentially refer to the whole panoply of artifacts, and possibly techniques, used in education. What goes along with this tendency to see the term *technology* only applying to *current* technology is the tendency to forget that: (1) what is no longer cutting edge was once considered ‘technology’; (2) sometimes these technologies, including the now-innocuous ballpoint pen, were once subjected to criticisms not unlike computers are today.

My principal focus here is the nature of this debate about educational technology, especially as it has appeared in philosophy generally and within the philosophy of technology specifically. Although most of my analysis focuses on philosophical critics of educational technology, this should not be seen as endorsement of educational technology. Instead, by carefully examining the logic of the critics’ arguments, I intend to demonstrate a serious flaw that appears in some of these arguments. Through arguing this, my hope is to provide a more productive strategy for understanding the relationship between education and technology. To this end, I begin with a brief survey of some of several philosophical analyses of educational technology. I then turn to Plato’s rather prescient critique of writing in the *Phaedrus* in order to show both the common root of many of these philosophical criticisms and a basic mistake that begins there. The concluding section then develops a pragmatic alternative for thinking about the relationship between education and technology in a more productive way.

Computers, Simulation, Education

Much of the recent philosophical discussion of educational technology tends to work in a similar vein as Friesen, criticizing the use of computers, especially the simulated nature of using computers in the classroom. One common criticism is that working with computers is so highly mediated that it cannot replicate the sort of experience that education requires to completely convey something to students. We can see this in *Holding on to Reality* where Albert Borgmann directly turns to education technology. While the terms he deployed in *Crossing the Postmodern Divide* are absent, his critique clearly builds on those lessons. Before the arrival of computers and other information technologies, what Borgmann refers to as “technological information,” education was accomplished through apprenticeship and lecture. Vestiges of these remain, inasmuch as they are simulated by educational technologies, but the depth and richness is lacking. The expansion of online education, in particular in colleges and universities results, in “The student becom[ing] the sovereign who can choose the material, the method of presentation, and the time and place of studying” (Borgmann 1999: 204–205). Under these conditions, education comes to resemble a grocery store. Students browse the aisles, selecting what looks interesting, while “Teachers are like butchers and bakers in the back who produce and package the

merchandise, like the clerks who keep the shelves full and up-to-date, and like the store managers who are ready to help and advice" (Borgmann 1999: 205). This marks a profound change in education. The goal of apprenticeship was the production of an artisan, while lecture aimed at producing scholars. Now education is increasingly treated as a mere commodity for teachers to manufacture and students to consume. This is a direct result of digitization because digitization allows for the easy transmission of information. By reducing texts, recordings, images, etc., to binary code, this information can be stored and retrieved in ways that are quite challenging for older media. One immediate implication of this is a change in the skill set students develop. On the one hand, the memorization of knowledge falls away (because it is all easily accessible via computers and other devices). On the other: "What the student needs are higher order skills, learning how to learn, finding whatever information is needed, and solving problems generally" (Borgmann 1999: 206). This makes the student and the computer somewhat analogous, in that both are oriented towards information retrieval and processing. What is lacking here is bodily engagement with the world. The world of computers in education lacks the sort of resistance that the world provides. This sort of friction helps to inculcate discipline, which is necessary in order to work with the ways that the world resists us like Friesen described. Without these sorts of richer engagements with the world, transforming information into knowledge and wisdom becomes less likely. Rather, we deal with hyperreal simulations of knowledge.

A different line of criticism comes from Hubert Dreyfus's "Anonymity Versus Commitment: The Dangers of Education on the Internet." Here Dreyfus argues that the internet is incapable of generating the sort of commitment by students necessary for meaningful learning. While the internet might be useful for some sorts of learning, following Kierkegaard, he argues that it is vital to possess an unconditional commitment in order to get as much as possible from education, specifically in order to live a meaningful life. Dreyfus works with Søren Kierkegaard's three spheres of existence (the aesthetic, the ethical, and the religious) to explore the sort of engagement that the internet does and does not allow for. The aesthetic involves approaching websites as a "source of boundless enjoyment [where] The only qualitative distinction is between those sites that are interesting and those that are boring" (Dreyfus 1999: 17). The student merely meanders from one webpage to the next in search of something that is "cool," i.e., will alleviate their boredom. For Kierkegaard and Dreyfus, this is not a place where one should rest one's existence because there is no commitment to anything beyond finding something fascinating—and as Dreyfus reminds us, just because it is fascinating does not mean that it is important, like spending hours looking at memes. The next sphere of existence is the ethical, which involves conditional commitment because "one has a stable identity and one is committed to involved action" (Dreyfus 1999: 17). This involves the adoption of a perspective on information through which one engages with it and the internet. This perspective allows one to sort out what is trivial from what is valuable, thus staving off the nihilism inherent in the aesthetic sphere. The problem with the ethical arises when one realizes that there are an infinite number of perspectives that one can adopt: "Any choice I make does not get a grip on me so it can always be revoked. It must be constantly reconfirmed by a new choice to take the previous one seriously" (Dreyfus 1999: 18). The only way Kierkegaard sees to avoid the nihilism of the aesthetic and the despair of the ethical is for a student to develop an unconditional commitment. This he refers to the religious sphere of existence. This sort of commitment cannot be revoked without serious harm to the person, because they build their world around the commitment. Unconditional commitments determine what is trivial and not, worth knowing or not, without the need to constantly recommit, as in the ethical sphere. These commitments are dangerous because they are risky. If what one is unconditionally committed to is removed or fails to be obtained, this does genuine harm to the committed because of how their existence is driven by the commitment. Yet, such commitments are necessary according to Dreyfus because "There is

no way to have a meaningful life and to develop particular skills and the skill of being a good human being without taking risks” (Dreyfus 1999: 19). If there is no risk, no possibility of genuine failure, then a meaningful life is impossible because it is this danger that allows for the richness of human existence.

Dreyfus’s central criticism of the use of computers in education is that they are incapable of supporting unconditional commitments. Computers do allow for the easy, if not all-too-easy, access to a vast amount of information. A simple web search produces a seemingly endless set of results, and most people’s use of the web tends to fit into Dreyfus’s claim that they flitter about from one page to the next to the next. The typical design of most webpages encourages this sort of interaction with them. In terms of education, this allows students only to access ‘information.’ This can be quite useful, but only if wedded to a commitment. If the student remains in the aesthetic sphere, all that can be found is information, with no means of separating the important from the rest. Put differently, the student finds a nigh endless amount of information, but never knowledge. In order to make this change from information to knowledge, the student must move to the ethical sphere. By adopting a perspective, the student is now in a position where they can distinguish between different bits of information. Furthermore the student can now focus in through the perspective and begin to acquire the necessary skills. While activity within the ethical sphere is more difficult via computers, Dreyfus holds that the internet can support these perspectives. It is more difficult because the internet appears to be oriented towards the aesthetic sphere, yet with effort the student can find ways through the maze of information to develop and utilize perspectives. However, only the aesthetic and ethical spheres can be supported on computers. The problem is this:

the choice of perspective that was supposed to turn the glut of information into knowledge and provide the involvement necessary for skill acquisition only adds to the possibilities, and one ends up in what Kierkegaard calls the despair of the ethical.

(Dreyfus 1999: 19)

The internet is incapable of allowing a student from avoiding this despair. While the internet can expose the student to a wide variety of perspectives, given the ease at which the student can switch from one perspective to another online, the internet only reinforces the student’s view that the choice of any perspective can always be revoked in favor of another. Here, the simulative aspect of computers reenters. Computers simulate experiences and engagements—but what makes these mere simulations is that there is no risk. If one fails using a computer, then one simply restarts. In which case, it is necessary to go beyond computers and the internet in order for students to explore spaces of unconditional commitments where risk is irreducible.

Educational Technologies as a *Pharmakon*

Many of the tropes and motifs that show up in the work of education technology critics first appear in Plato’s dialogue, the *Phaedrus*. What is of particular importance in our context is the rather strange conclusion of the dialogue where Socrates presents a myth about the origins of writing and then turns the discussion to the inadequacy of the written word for education. In this dialogue we find criticisms of writing echo those of contemporary skeptics of the educational use of ICTs. Plato argues that writing only creates the appearance of knowing, rather than genuine knowledge, much like Borgmann and Friesen, who criticize the use of simulations in the classroom. Concerns are also voiced about whether writing allows for remembering of things, or at best a shallow remembering, which is also echoed in Borgmann. Additionally, Dreyfus’s theme that computers do not allow for the sort of commitment necessary for living a fulfilling

life, which shows the limits of cultivating wisdom through ICTs, also appears here. Lastly, though going beyond Plato's intention, there are points that can be remolded into concerns about the commercialization of educational technology. This is not to imply a direct causal link between Plato and more recent critics, but rather that at least the general spirit, and some of the specific points, can be found in Plato. For this reason, it is worth spending a bit of time rereading the *Phaedrus* in in order understand a very basic, and powerful, articulation of the problems with educational technology and then map a general strategy for avoiding the problems this style of argument encounters.

The critique begins with Socrates telling a story of the god Theuth introducing the technology of writing to Egypt. The god brings many arts to the king of Thebes and the Egyptian gods, Thamus, and explains why they should be made available to people. Theuth describes the importance of writing as: "something that, once learned, will make the Egyptians wiser and will improve their memory; I have discovered a potion [*pharmakon*] for memory and for wisdom" (274e). Thamus responds to the god's claim, first by saying that Theuth is too close to the technology to objectively judge its merits, and then continues by stating:

it will introduce forgetfulness into the soul of those who learn it: they will not practice using their memory because they will put their trust in writing, which is external and depends on signs that belong to others, instead of trying to remember from the inside, completely on their own. You have not discovered a potion for remembering, but for reminding; you provide your students with the appearance of wisdom, not with its reality. Your invention will enable them to hear many things without being properly taught, and they will imagine that they have come to know much while for the most part they will know nothing. And they will be de difficult to get along with, since they will merely appear to be wise instead of really being so.

(275a–275b)

This amounts to Theuth inventing a shortcut to wisdom, a potion that creates the effects of wisdom without the necessary work to attain it. Given the nature of wisdom, and of knowledge, for Plato, there can be no such thing. As Socrates makes quite clear in his story, the best that writing can do is offer the appearance of wisdom and not the genuine article. The crucial problem with writing is that it is outside of the individual. The signs that form writing systems are public, owned by everyone and perhaps no one. In order to properly educate an individual, it is necessary for education to engage directly with the soul, which requires a dialectical relationship between student and teacher. Writing, in a significant sense for Plato, only appears to be "living." The trouble with words is that: "You'd think they were speaking as if they had some understanding but if you question anything that has been said because you want to learn more, it continues to signify just that very same thing forever" (275d). The fixity of the written words prevents the reader from truly engaging the content. This is why writing is at best a tool for reminding, as opposed to memory, since it only prompts its reader's soul to call up the memory form within themselves.

The spoken word avoids these troubles due to its intimate connection to the soul. Unlike the unchanging written word, the following can be said of speech: "It is a discourse that is written down, with knowledge, in the soul of the listener; it can defend itself, and it knows for whom it should speak and for whom it should remain silent" (276a). Although speech is "written down . . . in the soul," speech possesses a more dynamic character than writing. Writing always remains the same. Via speech, the teacher can tailor the discourse to the audience. If something requires defense, the speaker can do so. If something else is too difficult for the listener, the speaker should know not to bring it up. The written word runs roughshod over this because it

lacks the responsiveness of speech. Writing can say only what the writer puts down, at best. The writing might say too much and be too difficult for the reader, and it quite likely may say too little since it is only an image, like a painted picture as compared to the thing painted. For this reason, writing fails as an educational tool because it cannot convey knowledge and wisdom. Wisdom must grow up from within the individual, not relying on something outside of them. In the case of speech, the teacher facilitates the process of acquiring wisdom by dialectically engaging the student. The primary trouble with writing is that it fails to adequately replicate this sort of engagement. The teacher must be careful about what is to be taught to the student:

The dialectician chooses the proper soul and plants and sows within it discourse accompanied by knowledge—discourse capable of helping itself as well as the man who planted it, which is not barren but produces a seed from which more discourse grows in the character of others.

(276e)

Speech allows for this sort of dynamic relationship, while the written word remains fixed, dead, unable to respond to engagement. The written word cannot replace speech in education because it lacks the capacity to “write” knowledge and wisdom on the student’s soul. At the very best it might be an aid for remembrance. At worst, writing undermines the actual purpose of education.

The split between speech and writing is not merely that one is an effective technology and the other is not. Rather, *writing is technological and speech is not*, for Plato. Speech is immediately connected to the soul. He describes writing as having to properties mark it as fundamentally different, alien, from speech, marks that indicate its technological nature. In the myth, Theuth describes writing as a “potion,” which on its own hints at a connection to *techne* and technology. The Greek word that Alexander Nehamas and Paul Woodruff translated as “potion” is *pharmakon*. As Jacques Derrida analyzes it in “Plato’s Pharmacy,” the word *pharmakon* overflows with meanings: it can be translated as “permitted the rendering of the same word by ‘remedy,’ ‘recipe,’ ‘poison,’ ‘drug,’ ‘philter,’ etc.” (Derrida 1981: 71). All of these translations of *pharmakon* point to writing’s technological origins, because whether one translates it as “potion,” “remedy,” or “poison,” these are all human contrivances, inventions. Whether we understand *pharmakon* as either a medicine or a poison, or do as Derrida argues and understand it as both, it is the product of human activity. Medicines per se do not appear in the natural world: they must be produced by humans. So too with poisons that humans might use. This explains why: “the god-the-king [Thamus] nonetheless experiences the *pharmakon* as a product, an *ergon*, which is not his own, which comes to him from outside but also from below” (Derrida 1981: 76). Writing is an invention, a technology, brought to him by Theuth for his consideration—and, as we saw earlier, it is a product that Thamus judges rather harshly.

Education as Always Technological

This Platonic account of education, technology, and the relation between them is deeply problematic. To make this case, I begin with Plato’s distinction between speech and writing in education and then will generalize. The crucial assumption Plato makes is that speech is a natural extension of the soul, while writing is outside of it. For this reason, if education is in fact the cultivation of the soul’s wisdom, then speech will always be preferable to writing in education. This relies on a dualism between the natural and the artificial, wherein the natural is seen as inherently superior. Returning to Derrida’s analysis of the *pharmakon* in Plato, he notes that it: “goes against natural life: not only life unaffected by any illness, but even sick life, or rather the life of the sickness. For Plato believes in the natural life and normal development, so to speak,

of disease” (Derrida 1981: 100). The *pharmakon*, translated into English either as “medicine” or “poison,” is unnatural. If it is medicine, then it disturbs the natural course of an illness (see *Timaeus* 89b). If it is poison, then it disturbs the natural course of the living organism. In this way, writing is inherently disruptive because of its artificiality.

Regardless of whether the natural/artificial dualism can be maintained, Plato misuses it in this context. The argument that speech is natural, merely an extension of the soul, obscures the fact that all language, speech or writing, is a human invention. John Dewey puts the point this way:

But at every point appliances and application, utensils and uses, are bound up with directions, suggestions and records made possible by speech; what has been said about the role of tools is subject to a condition supplied by language, *the tool of tools*.

(Dewey 1958: 168 emphasis added)

Dewey does not draw clear lines between writing and speech like either Plato or Derrida. Rather, he subsumes both under the heading of “language” because they both are means of communication—and they are both tools; not just any tool, but the *tool of tools*. Language is the tool that allows for much human activity, technological or otherwise. Language proves to be a remarkably effective means for conveying information, like Dewey’s reference to “directions, suggestions, and records,” in order to transform the world through human action. Larry Hickman, who builds his approach to the philosophy of technology out of Dewey’s writings, defines technology in the following way: “Technology in its most robust sense, then, involves the invention, development, and cognitive deployment of tools and other artifacts, brought to bear on raw materials and intermediate stock parts, with a view to the resolution of perceived problems” (Hickman 2001: 12 emphasis removed). Between Dewey and Hickman, we begin to see that language is one, very important, tool that humans use in order to solve problems. This is not to dissolve the line between speech and writing entirely. Clearly, each possesses different properties. For example: there is a permanence to the written word that the spoken lacks, while Plato is correct that, at least in certain contexts, speech allows for responsive dialogue that writing cannot replicate (at least easily). That speech and writing possess different properties does not make one more ‘natural’ and the other more ‘artificial.’ Rather, this is a lot like hardware: a hammer and a saw, for example. Both are tools, but each possesses functions the other lacks. So too with writing and speech: each allows for different ways of communicating, but both share the property of being a tool for communication. From this follows the conclusion that Plato’s dualism between speech and writing cannot hold. In this Deweyian account, both are forms of language. While possessing different properties, neither has a claim to being more ‘natural’ than the other. This then dissolves the connection between speech and education, because it was speech’s ‘naturalness’ for Plato allowed for its direct connection to the soul.

This Deweyian line of reasoning points to a different understanding of the relationship between education and technology. Instead of the Platonic vision that holds that technology is fundamentally outside of technology, the suggestion here is that they are always thoroughly interconnected—so much so that while there might be technology without education, there can never be education without technology. In light of the previous comments about technology, especially from Dewey and Hickman, the claim that education is intimately linked to technology makes a certain sense. If technology involves the deliberate transformation of materials to solve perceived problems, then education qualifies as a technology in a broad sense. Here, there are two different sorts of materials at work. First is the material that undergoes transformation so that the student might learn it. Education rarely works by merely transferring data from one

place to another. Rather, we can agree with Plato that the teacher must take the information or knowledge they intend to convey and rework it in such a way that it is intelligible to their students. One goal of education can be to give students the same understanding of a subject that the teacher possesses—but in order to be successful, since the students do not possess this knowledge as the teacher does, it is incumbent on the teacher to reshape it in order for the student to understand the subject in question in a way like the teacher does (or at least be put on the path to achieving this). In this case the technology is the set of techniques deployed for transforming the subject matter. A second goal is that the students themselves should be understood as a material as well. Clearly, students are a material of a very peculiar kind. Rather than being information, knowledge, or wisdom, they are living organisms with interests, desires, and wills of their own (Garrison 1997). While any material can potentially resist attempts to shape it, students are different from materials like metal or computer code in that they sometimes have the capability creatively to thwart their teachers. Students are distinctive, but materials for technology to operate upon nonetheless. In this way, education itself becomes a technology for the transformation of students. Rather than the Platonic vision that holds that education and technology are fundamentally at odds, this alternative proposes that education is always technological.

On the one hand, this view that education is inside of technology does help to disrupt certain Platonic assumptions about the relationship between the two that continue through the present. At work in the criticisms that both Borgmann and Dreyfus make of the contemporary use of computers in education is the Platonic assumption that either technology inherently poisons education or the more moderate stance that the latest technology is poisonous. As discussed earlier, both see computers as always failing to simulate more traditional methods adequately. While there is some difference here between Borgmann and Dreyfus, in that Borgmann's brief discussion of the preparation of scholars leaves open some space for learning from the written word, whereas Dreyfus more clearly harkens back the necessity of the Platonic, intimate relationship between teacher and student, both agree that the newest technology (i.e., computers) undermines the aims of education without any apparent qualification. By dissolving the line between technology and education, however, the argument that (the latest) technology is necessarily poisonous does not follow.

On the other hand, to say that education is always technological is infuriatingly vacuous or dangerous. It might be vacuous inasmuch as it seems to stop the slide into the Platonic account, but little more. It says only that technology and education are connected, but nothing about how or its significance. It is quite dangerous to read this as a blanket endorsement of all educational technology for two reasons. First, this simply inverts Plato's mistake. Second, it runs the risk of obscuring the pressures that drive the adoption of many contemporary educational technologies (specifically that the adoption is driven more by administrative cost-cutting measures, agreements between manufactures and educational institutions, undermining the labor power of teachers, generation of revenue, etc.), i.e., pressures that sometimes force the use of these technologies not based on their pedagogical merit. So, making the global claim that all educational technology is inherently beneficial is deeply dangerous. Instead of this move, the view that all education is inherently technological opens a space for exploring in a more sophisticated way about what is beneficial and harmful about using particular technologies in education.

The promise of the view proposed here can be seen by accepting that different technologies function in sometimes incommensurable ways. This reframes our understanding of educational technology in a non-Platonic way through dealing with particular examples of educational technologies. Beyond this vacuous statement, there are few global claims that can be made about the nature of educational technology. A more subtle approach is necessary, one that assesses

the value of a particular technology in light of educational goals, a contextual evaluation of specific educational technologies. The proper assessment of an educational technology must involve considerations of the educational goals, the capabilities of the teacher and students, the technology itself, and so on. For the same reason that global assessments of technologies fail, so to with educational goals. Instead, such questions must be asked within a particular context, thus limiting possible, acceptable answers. Following Hickman (1990) and Hans Joas (1996), for Dewey ends and means are reciprocally connected. What one takes as an end should be informed by the means available. For example, attaining certain goals might become prohibitively difficult because one lacks the means necessary. In such circumstances, this sort of end is otiose, because it cannot play a substantive role in practical action (it is not what Dewey refers to as an “end-in-view.”). The reverse is also true: what will count as a means will be informed by one’s ends. In the abstract, a seemingly infinite number of things could be taken as a means. Yet, within the context of practical action, a particular end-in-view will establish what are more or less promising means for realizing that goal.

Taking these elements of Dewey together, we find a more productive way to think about educational technology. This encourages us not to engage in a blanket condemnation, or embrace, of educational technology (either in the sense of all technology or just the latest) generally. Instead, the particulars matter: both the particulars of the technology and of the educational goals. Because ends and means are reciprocally linked, the assessment of a particular technology should be done only within the context of specific ends-in-view, while those technologies will in turn shape what goals educators set for themselves (or are imposed on them by administrators; this is somewhat problematic). Returning to Borgmann, Dreyfus, and Plato’s critiques of educational technology, they are surely correct that there is much to be critical of with the use of ICTs and other technology in education. The trouble with their arguments is that they overgeneralize, in effect fixing the ends of education and educators, then dismissing anything that does not fit neatly with this particular end. Yet, unless one stays at the level of vague and uninformative generalities, it is impossible to say what *the* specific end of education is—rather there are many ends, which depend on the context, the teacher, the students, the material, the problems to be solved, and so on. The pedagogical value of a particular technology must be determined with respect to these specific ends. In some circumstances, computers are disruptive to, or destructive of, education. In others, as Robert Rosenberger (2011) argues in response to Norm Freisen, and as James Petrik, Talgat Kilybayev, and Dinara Shormanbayeva (2014) argue in response to Dreyfus, computers can actually be quite beneficial in furthering particular educational goals. Rosenberger and Petrik, Kilybayev, and Shormanbayeva pay careful attention to context, being sure to note where particular ICTs work as means for attaining specific pedagogical ends and where they do not.

It is not only the responsibility of teachers to engage in this sort of reflection on pedagogical ends and means. Rather, everyone involved in education (teachers and students, as well as administrators, parents, and so on) should be involved because, in a point of agreement between Plato and Dewey, education is transactional (Garrison 1997). Those participating in the educational process must be aware of how they engage and affect the other participants. This means that everyone involved must think critically about the sorts of educational technologies. This points to another benefit of this view of education and technology: it fosters users of technologies to think about the technologies they use. Given the increasingly prominent role that certain sorts of ICTs (smartphones, tablets, computers, etc.) play in students’ lives (and many others as well, including their teachers), encouraging them to reflect on how they ICTs and what ICTs do them puts them in a better position to evaluate their own use of those technologies, both within education and beyond.

References

- Borgmann, A. (1992). *Crossing the Postmodern Divide*. Chicago: University of Chicago Press.
- Borgmann, A. (1999). *Holding on to Reality: The Nature of Information at the Turn of the Millennium*. Chicago: University of Chicago Press.
- Collins, A., and Halverson, R. (2009). *Rethinking Education in the Age of Technology: The Digital Revolution and Schooling in America*. New York: Teachers College Press.
- Derrida, J. (1981). *Disseminations*, trans. B. Johnson. Chicago: University of Chicago Press.
- Dewey, J. (1958). *Experience and Nature*. New York: Dover Publications.
- Dreyfus, H. (1999). Anonymity Versus Commitment: The Dangers of Education on the Internet. *Ethics and Information Technology*, 1, 15–21.
- Friesen, N. (2011). Dissection and Simulation: Brilliance and Transparency, or Encumbrance and Disruption? *Techne*, 15, 185–200.
- Garrison, J. (1997). *Dewey and Eros: Wisdom and Desire in the Art of Teaching*. New York: Teachers College Press.
- Hickman, L. (1990). *John Dewey's Pragmatic Technology*. Bloomington, IN: Indiana University Press.
- Hickman, L. (2001). *Philosophical Tools for a Technological Culture: Putting Pragmatism to Work*. Bloomington, IN: Indiana University Press.
- Joas, H. (1996). *The Creativity of Action*, trans. J. Gaines and P. Keast. Chicago: University of Chicago Press.
- Petrik, J., Kilybayev, T., and Shormanbayeva, D. (2014). The Internet, Identity and Intellectual Capital: A Response to Dreyfus's Critique of e-Learning. *Ethics and Information Technology*, 16, 275–284.
- Plato. (1997). *Phaedrus*, trans. A. Nehamas and P. Woodruff. In J. Cooper (ed.) *Complete Works*. Indianapolis: Hackett Publishing.
- Rosenberger, R. (2011). A Phenomenological Defense of Computer-Simulated Frog Dissection. *Techne*, 15, 215–228.

Religious Transcendence

Carl Mitcham

Can we transcend technology? Should we want to? Arguments have been made for and against, mostly from within various religious traditions. There are also arguments for technology itself as a means of transcendence. In philosophically oriented religious discussions of technology, the issue of transcendence may attract increased attention in the future.

Historically and phenomenologically, humans have regularly described their experience in terms of two levels, commonly denominated as lower or higher—sometimes as superficial or deep. The spatial metaphors suggest one level of experience is always present and dominant but less significant or meaningful, another is elusive but more meaningful. In some instances the always present and dominant is even described as connected with an inauthentic or false self and the effort to connect with the higher or deeper level as the cultivation of a true self or authenticity. Persons who have strongly realized the higher or deeper levels are often described as saints, sages, or mystics. Critical reflection on this particular duality in human affairs and its precise structure is a major topic in the history and philosophy of religions.

According to Mircea Eliade, a seminal figure in framing the scholarly study of religions, what distinguishes religion from other aspects of culture is the phenomenon of the sacred. The sacred is characterized by its opposition to the profane, secular, or worldly. Religion assumes or develops in culture the notion of the sacred. The sacred is singled out as an aspect of experience that is special or distinct from and rises above the everyday. The sacred can be an event in human life (birth, puberty, marriage, death), a natural place (sacred mountain), a structure (temple, church, mosque), a behavior (ritual), a teaching (doctrine), an artifact (bible, prayer wheel, icon), or a person (shaman, priest, monk). Another term that can be used to call attention to what is distinctive in this multiplicity is *transcendence*. What follows is thus a brief, episodic reflection on transcendence and its problematic character in relation to technology and the techno-lifeworld.

From the beginning, it is important to set aside a Western propensity to identify transcendence solely with its understanding in the Jewish, Christian, or Islamic traditions. In these Abrahamic religions transcendence originates with God, a wholly other supernatural reality, who creates the world as other than divine, and then takes an initiative to reveal himself or enter into this world. In the dharma and dao religions of Asia (Hinduism, Buddhism, Daoism, and Confucianism) transcendence arises with efforts by human beings to go beyond themselves or the world. In the dharma-dao religions transcendence happens, as it were, from the bottom up. The Abrahamic religions reject this as false transcendence and claim that the only true transcendence is top down, with a revelation from God. The other great category of religions, the so-called indigenous religious traditions, experience transcendence as a given in human experience. The sacred interpenetrates the profane; transcendence is ubiquitous, and does not require much in the way of special effort on the part of humans or any unique initiative on the part of the gods or God. The gods are already present everywhere in everyday life.

One basic question that bridges the philosophy of religion and philosophy of technology concerns whether and to what extent, with the historical emergence of the techno-lifeworld, something distinctive takes place with regard to transcendence as understood or experienced in these three types of religion. Broadly speaking, the transcendence-technology interaction can be negative, neutral, or positive. In fact, arguments and interpretations have been advanced for all three positions across different religions. A preliminary introduction to the transcendence-technology relationship may thus be sketched by selectively noting different perspectives.

Transcendence and Techno-Lifeworld in Indigenous Religions

The experience of the sacred as ubiquitous in indigenous religions is well-represented in traditional craft technics. Foods and their preparation, clothing, utensils, and tools all reflect transcendence through myths of their origins and ritualizations in practical uses. This interpenetration of the here-and-now of technics and the transcendent is fundamentally shattered by the coming of natural philosophy and subsequently of scientific technology. In the first instance, instead of the ways of all things as originating with heroic figures or gods, natural philosophy distinguishes between the ways of cultural things (languages) and phenomena independent of culture (things of nature). The differences between Greeks and Egyptians are cultural, but water is equally wet by nature in Greece and in Egypt. The efficacy of the transcendent in nature is replaced by natural phenomena themselves conceived as their own immanent realities; all that is called for is insight into the basic structures of such realities. Natural philosophy relegates myth to the phenomena of culture.

Then in a second instance, when the technics of culture are replaced by modern technologies grounded in science, there is an undermining of myth and tradition also in material culture. It is not a cultural hero in a primordial past who gives to a people their technics but inventors and industrialists who produce consumer goods that people must purchase with money earned by working in factories. The anthropological story of the Yir Yoront peoples of Australia exemplifies the result. Once their stone-age tools are replaced by steel axes, an integrated sacred-secular culture loses its way and falls apart. In this circumstance, technology is opposed to and destructive of transcendence.

Transcendence and Techno-Lifeworld in Abrahamic Religions

What Karl Jaspers termed the “Axial Age” of human history witnessed the independent but parallel emergence of philosophy in Greece and what are now the world religious traditions: Abrahamic in the West, dharma-dao in the East. Fundamental to the Axial Age is the introduction into culture of a question concerning what it means to be human. Prior to the Axial Age peoples understood their humanness as given by birth, place, and tradition. Socrates, the Hebrew prophets, Shakyamuni Buddha, and Kongzi (Confucius) all challenged humans to live differently than they might otherwise have lived, to transcend their cultures and themselves, by means of techniques of nonmaterial culture. Socrates introduced the *elenchus* into thinking; the prophets asked Hebrews to turn their hearts to God; Buddha offered a path of meditation to transcend the cycles of karma; Kongzi initiated a tradition of literati self-cultivation.

It is possible to view this emergence as made possible by the flourishing of material culture technics in the domestication of plants and animals and the corresponding appearance of the great civilizations of Egypt, Mesopotamia, India, and China: technics preparing the way for a new type of culture. One provocative theory under investigation with technologically enhanced social scientific methods postulates the existence of “big gods” as required to mediate prosocial behavior in large groups that can no longer rely on kinship relationships (Norenzayan 2013).

These big gods ask humans to transcend their limited tribal ways in favor of relationships less embedded in biology. Precisely because of the disruption of natural transcendence in indigenous religions, a new transcendence is institutionalized into the social order using omnipotent and omniscient gods.

Within such cultures of god-based transcendence, historians of religion further distinguish primary and secondary formations. Primary religions emerge in continuity with indigenous traditions; secondary religions emerge in opposition to primary religions. Hinduism, Daoism, and the literati tradition commonly called Confucianism all emerged as refinements of previous indigenous traditions. Judaism came to be in opposition to an existing big god religion in Egypt; Buddhism initially as a sect within Hinduism; Christianity in rebellion against Judaism; Islam in violent rejection of multiple indigenous traditions. In the ambience of such cultural oppositions, it is likely that other oppositions will also be nourished.

Medieval historian Lynn White Jr. (1978) argues that such was indeed the case, especially within the Christian tradition. In the course of the cultural development of the Christian rebellion against Judaism, Christians began to think of God as having assigned to them the task of world transformation: “God said unto them, Be fruitful, and multiply, and replenish the earth, and subdue it: and have dominion over the fish of the sea, and over the fowl of the air, and over every living thing that moveth upon the earth” (Genesis 1:28). In Francis Bacon’s modernist reformulation, Christians should recognize the historical power and human benefit of such inventions as the printing press, gunpowder, and the compass, and promote further inventions in a spirit of *caritas*. All such Christian investments in technological progress are, paradoxically, associated with a reification of the two-level experience at the core of religious experience. The absolute transcendence of the Abrahamic God and belief in a strong life after death, especially when allied with a Platonic two-world ontology, justifies doing almost anything in this life, provided it is understood to be sanctioned by divine command. There is no need for worldly prudence.

A further paradox of this Christian endorsement of technology is that it has tended to undercut Christian practice and belief. As Europe developed technologically, church attendance waned. Although the decline has been less in the United States, American church affiliation remains strongly correlated with fundamentalisms that violently oppose aspects of scientific and technological progress (such as theories of evolution and technologies of abortion). Maintenance in the techno-lifeworld of belief in transcendence through Christian faith and practice becomes problematic.

Of course, fundamentalist and liberal Christians alike are not adverse to using technological media to promote their faiths and to criticize the techno-lifeworld for what appear to secular majorities as prudential accommodations: governmental regulations of technological risks and global collaborations with peoples of different faiths and practices. Such religious media propaganda nevertheless continues to exhibit its own rhetorical violence, with the case of Islam being particularly revealing. Egyptologist Jan Assmann (2009) suggests that all Abrahamic religions (as well as their secular progeny) partake of this heritage, which is grounded in a monotheism that rejects as false both atheism or polytheism.

Transcendence and Techno-Lifeworld in Dharma-Dao Religions

The situation with regard to (atheistic and polytheistic) dharma-dao religions in the techno-lifeworld is different than the Abrahamic religions. The reason is twofold: the religions themselves never arose in association with the level of violence manifested in the Abrahamic religions and technology itself did not arise within a dharma-dao cultural ambience.

At the same time, the need to import technology from the West in the face of Western imperialism has obviously been a challenge. Yet precisely insofar as modern technology was forced

upon it, the East has found it possible to see technology as less essential to culture than in the West. In the West, Hegel and Marx are no more than the most philosophical defenders of the idea that technological engagements with nature are fundamental to what it means to be human and to historical development. Benjamin Franklin's definition of man as *homo faber* and Samuel Florman's "existential pleasures of engineering" popularize a common Western consensus. Technology is not to be transcended but affirmed.

By contrast, Chinese thinkers have often adopted a Buddhist essence-function distinction. During the late Qing dynasty, Confucian reformers in the Self-Strengthening Movement summarized their position with the phrase "Chinese learning for essence, Western learning for application." The essence of Chinese culture could be maintained while importing Western technology for practical application and economic development. In a sense, China, Japan, India, and other Asian countries, while recognizing its utility, do not take modern technology quite as seriously as countries in the West. Indeed, until forced to do so by Western imperialism, China resisted modern technology in favor of alternative traditions of this-worldly engagements that over more than a thousand years, as Joseph Needham has documented, gave China a civilization more technically advanced than any other.

Various explanations have been offered for the alternative trajectories of technical change within dharma-dao traditions. One candidate is what contemporary philosopher Li Zehou calls a "one-world ontology" that pursues a transcendence of aesthetic pleasure. To simplify, in China transcendence has traditionally been constituted through integration with that immanent reality known as dharma or dao, which at the same time is not subject to strict conceptual articulation. As the opening of the *Laozi* has it, "The dao that can be dao-ed is not the unchanging dao . . . having no name, it is the originator of heaven and earth." In place of the nature/culture distinction characteristic of the West, humans and their culture are seen as positioned between heaven (sky or all above) and earth (everything below) and ordered toward an aesthetic life of cosmic harmony. Any two-level differentiation is internal to the whole.

Buddhism likewise preserves a one-world ontology. From a Buddhist perspective, the deeper level of transcendence is attained simply by being fully present to the superficial level, the *samsara* of existence. As Vietnamese Buddhist practitioner Thich Nhat Hanh has put it, washing the dishes to clean up after dinner is to fail to be fully present, to wash the dishes to wash the dishes. Meditation is a training to pay attention to what is at hand, to see it for what it is, not be distracted by the way everydayness presents itself as something to get over with and thereby becomes *samsara*: "Before enlightenment, chop wood, carry water. After enlightenment, chop wood, carry water." The technologies of the contemporary lifeworld, precisely insofar as they present themselves as instrumentalities for the achievement of other (often personal and ephemeral) ends, become especially distracting forms of *samsara*.

Nevertheless, Buddhists commonly see the techniques of meditative practice that seek to attend more deeply to and thereby transcend everydayness—including the *samsara* of philosophy and religion—as available to be applied anew in the techno-lifeworld. As one self-described "Buddhist geek" notes, there is the special "challenge of maintaining presence in the midst of an always-on digital environment" (Horn 2015). Defending the efficacy of traditional meditation techniques, he also postulates the potential of new contemplative promoting apps that he terms "technodelics."

Secular Transcendence

Even without meditation, many nonreligious persons find themselves quite at home in techno-everydayness and claim to find in an identification with it a kind of authenticity. While technology has deconstructed traditional forms of the sacred (medicalizing the events of birth and death, mining sacred mountains), the laboratory has become its own kind of sacred space and

technoscientific innovation a sacred practice. Technoscientific workers at the cutting edge of knowledge production and innovation easily find themselves transcending the everydayness of personal relationships (making money, eating and sleeping, getting along with friends or spouse). Ray Kurzweil even projects an ultimate transcendence of death by technological means: anti-aging biomedical therapies, the uploading of the mind into computers, and AI as a new stage in cosmic evolution. Among others, however, Stephen Hawking warns about the potential dangers of AI transcendence. Not only the good, but evil too—as in warfare and crime—can produce experiences of transcendence.

Speculative Conclusion

Transcendence as a philosophical term has scholastic origins. The Aristotelian transcendentals are categories of being common to all beings: *unum*, *verum*, *bonum*, et al. (There are disagreements about the precise number of these transcendentals.) In medieval theology, God is beyond even the transcendentals: God is absolutely transcendent, not immanent to experience, only self-revealed. For Kant, the transcendentals are those a priori concepts that condition all human knowing and are immanent within intuition, whether sensory or intellectual. In philosophy, transcendence is thus opposed to immanence; and in modern philosophy, which emphasizes the primacy of sensory experience, it has become relegated to existential psychology. For Jaspers, Martin Heidegger, and Jean-Paul Sartre, humans must struggle to transcend their inauthentic selves that are dominated by the everydayness of social relationships. Insofar as our everydayness has become fused with technology, this naturally becomes a struggle to transcend the techno-lifeworld, and represents a prolongation of the Axial Age infusion into culture.

The contemporary world is nevertheless also giving birth to a New Axial Age. The original Axial Age question—What kind of humans should we be?—is now complemented by a new one. Sponsored by technological increases in power, impact, and design—from the macro (nuclear energy) through the meso (chemical and biological) to the micro (nanoscale) levels—we are increasingly impelled to ask: What kind of planet should we construct?

Responses to the original question could be diversified across noninteracting cultures. The Jewish view of being truly human as responding in obedience to an absolutely transcendent God was not the same as the Buddhist view of meditative detachment from karma—and that was OK, because Jews lived on one side of the planet, Buddhists on another. As the world has become fraught with global social interactions, the need for some kinds of accommodation have emerged. But as the world becomes an artifact of globalized human engineering, the new question of what kind of world should we live in demands more than the development of cultural tolerance. It requires a globally agreed-upon answer about the want to make of our common planet.

Given technological globalization, we cannot help but long for the end of history. Forget economics and politics: there cannot be diversity in the ways we manage nuclear weapons, the stratospheric ozone hole, or a planetary climate. Geoengineering, whether undertaken unintentionally (by allowing the continual production of greenhouse gases in the name of economic development) or intentionally (launching aerosols into the stratosphere to cool the planet) affects everyone. The clash of civilizations is not just an issue of international relations, but of planetary design and management that demands globalized agreement.

Is it possible to transcend the trajectory of technologies—trajectories pointing toward, for example, a global design and management of climate, universal informatization and surveillance, pervasive genetic engineering, and artificial intelligence—and to think reflectively and critically about what we are doing? Will it be possible to transcend the disagreements that appear to be only too inherent in the lower or superficial realms of politics, as manifested in the

challenges of globalized social interactions, when political agreement must be reached about the planet as a whole? Is the New Axial Age question even in principle answerable?

References

- Assmann, J. (2009). *The Price of Monotheism*. Stanford, CA: Stanford University Press.
- Eliade, M. (1959). *The Sacred and the Profane: The Nature of Religion*, trans. W. R. Trask. New York: Harper.
- Florman, S. C. (1976). *The Existential Pleasures of Engineering*. New York: St. Martin's Press.
- Horn, V. (2015). Buddhism in the Internet Age. *Inquiring Mind*, 31(2), Spring.
- Kurzweil, R. (2006). *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking.
- Needham, J., et al. (1954–2008). *Science and Civilization in China*. Cambridge: Cambridge University Press.
- Norenzayan, A. (2011). *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton, NJ: Princeton University Press.
- White, Jr., L. (1978). *Medieval Religion and Technology*. Berkeley, CA: University of California Press.

Agency and Citizenship in a Technological Society

Andrew Feenberg

Introduction

Langdon Winner argues that technology resembles legislation in its power to shape our everyday existence (Winner 1992). If the comparison holds, it seems there is a *prima facie* case for recognizing a right to citizenship in the technical domain. But this conclusion is contested by what I will call *technocratic ideology*. This ideology claims that public agency conflicts with the role of specialized technical experts in fields such as medicine and engineering. Only experts should have the right to intervene in such cases. Some advocates of this view extend the claims of expertise from technology to society as a whole: technocracy as the rule of experts.

In this chapter I argue that technocratic ideology fails as an account of technology, not to mention as a political program. The issue is the legitimate role of the public in technical decision making. To begin with I introduce considerations on the nature of citizenship, followed by a discussion of the technocratic position and its gradual collapse in the face of social movements such as the environmental movement. To conclude, I introduce a philosophy of technology that supports the claims of technical citizenship.

Technical Citizenship

Citizens have rights: for example, the right to express their opinions. But if a citizen's expression of opinion remains a purely private matter, with no possible impact on the world, we do not consider them to be fully exercising their citizenship. Even the right to vote is insufficient to qualify individuals as real citizens unless the elections are free. Citizenship requires more than rights. It also involves something we call *political agency*. What do we mean by this concept of agency?

Some definitions of agency simply equate it with the capacity to engage in intentional action, but a restricted definition seems to conform better with the everyday distinction we make between actions that emphatically implicate the subject and others that are more or less thoughtless or impersonal. In this restricted sense agency implies three conditions: knowledge, power, and an appropriate occasion.

The connection between power and knowledge is clear: individuals who do not understand the implications of their own actions may do themselves more harm than good. This is not what we usually mean by agency. The concept implies more than the ability to act. To qualify as agency, the action must have the power to achieve goals beneficial to the subject, or at least reasonably believed to be such. For example, children lack agency in the medical sphere, as do voters successfully manipulated by propaganda. The agency and hence the responsibility or culpability lies with the parents or the propagandists.

The role of power in agency is qualified by the fact that we act effectively in many circumstances in which we have no real agency. This is the case where a universal consensus prevails,

as when the subject's actions conform to a cultural norm. Thus using a knife and fork at table is not properly described as agency except in cases of disability and age. Nor is agency relevant where rationality dictates uncontested solutions to problems, as in the case of simple arithmetic.

We reserve *agency* for domains in which action is both personal and informed, and in which it is appropriately so. Politics is the prime example, and we call *agency* in this domain citizenship. *Citizen agency* is the legitimate right and power to influence political events on the basis of personal opinion and knowledge.

This leads to the questions that I will address in this chapter: is there something we could call *technical citizenship*? If there is, what is its relation to political citizenship? The answer to these questions depends on the understanding of the nature of technical problems. From the technocratic standpoint, they are seen as similar to mathematical or scientific problems. On these terms technical problems ought to be solved on the basis of evidence acquired and analyzed by rational procedures like those of natural science and mechanics. In the case of technology, that truth involves the design of artifacts and systems. This is a matter of fact rather than opinion. Politics has no place here. From this technocratic standpoint, disagreement with the facts asserted by the technical experts is simply irrational.

There is something to be said for this view; no one wants a referendum on bridge design. But this is a straw man: the technocratic argument rests on a hidden assumption, namely, that in their domain technical experts know *everything* relevant and rational that can be known. Thus the real question is: do the users and victims of technology know anything worth knowing that technical experts do not already know? This formulation reveals the problem with technocracy. There are obvious obscurities and points of ignorance in technical disciplines just as there are in every other type of knowledge. There are interests at stake, there are traditions, and there are of course errors. Here I will focus on two issues, (1) the limitations of specializations and (2) the delays in error correction.

By their nature, technical disciplines leave out much important information. In the real world everything is connected in a concrete whole, but specializations isolate and separate out a particular cross-section of reality for analytical treatment. Technical disciplines are based on a simplified understanding of their objects, to which correspond reductive and simplifying practical procedures in the making of technical devices. Reduced to raw materials and disconnected from their natural background, the materials incorporated into technologies can have unanticipated side effects that become fatally significant as they mediate more and more of social life. Eventually these side effects cause such destruction and disease that ordinary people are affected and protest.

Despite our reasonable everyday confidence in technical experts, they do fail from time to time. Bridges have fallen because the engineers who built them did not anticipate the effect of local winds. Everything was eventually explained through combining several disciplines, but only after the failure of the bridge. Similar problems arise in relation to worker health and safety. Often the hazards associated with chemicals used in industry are identified only after the fact, when the workers report symptoms to a doctor.

Because there is no metadiscipline able to predict the need for multiple forms of disciplinary knowledge to deal with unanticipated problems, feedback often comes first from experience. In many cases, those who live with the technology—those who build it or use it or suffer its side effects—bring its limitations to the attention of the experts.

They may of course be wrong, and when they are further investigation discredits protests, which gradually die down. But where they are right and identify a real problem, the process of correcting errors in technical disciplines and practices is complex and time consuming. Disputes can go on for years due to dogmatic attachment to the current state of a discipline or the will of powerful organizations to protect their interests regardless of consequences. Eventually

the lessons of experience are incorporated into the flawed discipline on technically valid terms. Designs are modified to reflect a more realistic understanding of nature's complexity. This overall dynamic leads to awareness of the hybrid character of technology and a weakening of technocratic and determinist ideology. Predictably, technical politics will become part of mainstream political life as this process unfolds.

The commonplace notion that politics concerns disagreement over values or matters of opinion rather than facts turns out on these occasions to be only partially correct. Of course this is true of many political disagreements. But in the case of technological disagreements, facts are in dispute, and the disputes look very much like political ones. In this context, appeals to expert authority are not authoritative. All parties to the dispute must tolerate disagreement—but this is just what we expect of citizens.

The narrowness of technical specialization has a second consequence for citizen agency, as illustrated by events in domains such as information technology. Introduced in the context of military and business enterprise, these technologies have been colonized by users in pursuit of opportunities to communicate. The opportunities opened by information systems have a role parallel to that of side effects in environmentalism, revealing complex potentials of the systems unsuspected by their original designers. These potentials are benign rather than threatening and deserving of independent development. They enable new forms of sociability and multiply creative possibilities for ordinary people. The democratic implications of these technologies emerge as resistance grows to commercial exploitation and political suppression.

There is room for experiential knowledge concerning technology. This is an informal knowledge from below based on the everyday experience of ordinary people. It is occasioned by harms or unexploited potentials of technology that have been ignored by the experts but that users identify on the basis of their experience and imagination. The chief examples of these two categories are the harms to health of industrial pollution and the communicative potentials of the internet.

These considerations on technical knowledge suggest that there may be a kind of citizenship in the technical domain. Recall the conditions of agency: knowledge, power, and an occasion. I have already suggested that ordinary people may have useful knowledge relevant to an appropriate occasion of some sort. Now I will consider whether they have the power to make changes.

The Technocratic Ambition

The notion that wise technical experts could rule better than kings or citizens has a long history that goes back at least as far as Francis Bacon's *New Atlantis*, but only in recent times has it taken the form of a popular ideology. This ideology first began to have wide appeal with the rise of large-scale technical systems such as the railroads and the electrical system. These giant macro-systems were efficiently regulated by small cadres of engineers and bureaucrats. They encompassed the total society and transformed its daily life (Wagner 1998: 230–235). Early criticism and utopian expectations soon gave way to acceptance. In the end, ordinary people did not expect to have agency in the railroad and the electrical systems. A new principle of authority came to be generally accepted.

These early large-scale technical systems were taken to represent the essence of technology and to point the way to a new form of rational society. Some thought they were on the path to utopia; for others these systems inspired dystopian despair. But the influence of technocratic ideology was rather limited until the 1950s and 1960s. At that time, technology and bureaucratic systems seemed to spread into every corner of the social world. The development of new programming and economic tools after World War II gave real plausibility to the technocratic

pretension to regulate society as a whole. The early 1960s was the highpoint of this ambition. Its defeat occurred in three stages. The consequences of these defeats are vast, but the changes they have wrought are simply taken for granted and the role of public participation is forgotten.

This is the case, for example, with the popular dystopian critique of centralized power in modern societies. A critique developed by a few social theorists in the 1950s entered popular culture in the following decade. Young rebels at first protested the limitation of democratic procedures to electoral politics, while in their everyday life the citizens were subordinated to management and administration at work, in dealings with medical institutions, government agencies, even unions and political parties. In the 1960s movements for political participation challenged the technocrats. The concept of ‘alienation,’ hitherto an obscure technical term in Hegelian and Marxist philosophy, became a popular slogan.

This was one of the prime inspirations of the American new left. The student movement called for “participatory democracy,” by which was meant general consultation and consensus decision making rather than top-down control. In France in 1968 a much more powerful movement with wider popular support demanded self-management in the economic and political institutions of the society (Feenberg and Freedman 2001). These protests engaged millions of people.

Demands for participation were relayed in the 1970s and 1980s by more focused movements. Against considerable resistance from business, environmentalists demanded regulation and alternative technologies. Demonstrations and public events such as Earth Day had a huge impact on public opinion and, indirectly, on technological design as well. Environmentalism quickly proved that public participation is neither impotent nor incompetent. If we are all aware of environmental issues today, it is largely owing to these social movements.

Movements in the medical sphere also had an impact. Although there have been recent setbacks, the 1970s saw major changes in childbirth procedures under pressure from women and women’s organizations (Michaels 2014).¹ A decade later, AIDS patients fought to transform the practice of experimental medicine (Epstein 1996). Another similar movement among AIDS activists in the 1980s started out with considerable conflict and distrust between patients and the scientific-medical community. Patients objected to restrictions on the distribution of experimental medicines and the design of clinical trials. The struggle eventually died down as the leaders of patient organizations were invited to advise scientists and physicians on a more humane organization of research. This lay intervention added a new ethical dimension to scientific practices. While the importance of informed consent achieved nominal recognition in the 1960s, this was the first time that clinical research officially recognized satisfaction of patient needs as rights. The changes were also cognitively significant since they made it easier to recruit human subjects and to ensure that they cooperated in supplying researchers with information.

The third stage of the process emerged with the internet in the 1990s and continues down to the present. The internet provided an example of technical potentials invisible to the experts but known to users who realized them through hacking and innovation. Human communication is the most significant of these potentials. It was not envisaged by those who originally created the internet to support time sharing on mainframe computers (Abbate 1999: 108).

With the internet, a new paradigm of the relation of human beings to machines entered the popular imagination. Where formerly large-scale technical macro-systems symbolized the conquest of society by technology, now the personal computer seemed to reinstate the agency of the individual in the technical sphere. As hackers and amateur innovators worked their magic on the internet, everyone was shown brilliant examples of a new kind of technical micropolitics that enhanced the established technical systems while subverting their original design (Lievrouw 2011).

These movements led to the decline of expert authority. Meanwhile, a new paradigm of the relation of human beings to machines entered the popular imagination. Where formerly

large-scale technical macro-systems symbolized the conquest of society by technology, now the personal computer seemed to reinstate the agency of the individual in the technical sphere.

With the expansion of the public sphere to include technology, new forms of technical agency have emerged. How significant this evolution will turn out to be is still in question. I believe we are at the beginning of a process that will eventually institutionalize public involvement in technical decision making. Naturally, this is not an unmixed blessing. The public makes mistakes too, as for example in the case of the rejection of vaccinations for childhood diseases. But every advance of democracy grants new powers to the 'unqualified.' Only after the individuals have obtained the power to participate do they engage the learning process that qualifies them to exercise it. So far, in any case, the public has not done so badly in technical matters. For every awful case such as the rejection of vaccinations one can find multiple examples of public action leading to significant technical improvements, such as the removal of lead from household paint or the introduction of e-mail in the shadow of the official uses of computer networks.

Democratic Rationalization

Critics of technology often argue that progress should have a moral as well as a material dimension. This was the promise of the Enlightenment, but it has been sadly disappointed. Citizenship in the technical sphere revives hope that it can someday be fulfilled. But is morality compatible with instrumental rationality? No doubt particularly egregious abuses may provoke legal or regulatory actions in response, but is this any more than a marginal corrective of a process of technical development that is essentially amoral?

This technocratic argument is based on premises that go back to the foundations of modern social science. Max Weber argued that progress results from the increasing role of calculation and control in modern societies. This is called *rationalization* in the sociological theory that derives from his work. Rationalization in Weber's sense refers exclusively to means; it concerns progress in instrumental rationality. According to Weber and his numerous followers (many of whom are not aware of the source of their views), modernity describes a society based on rational means. Weber also held that no comparable rationalization process determines ends. Many observers of modern societies agree that the overall goals of modernization are inherently nonrational, if not actually irrational. Weber concluded that bureaucracies informed by technical knowledge but without moral guidance would gradually establish an "iron cage" dictating the form of modern life (Weber 1958). The first generation of the Frankfurt School echoed this conclusion. Its theory of instrumental rationality went beyond Weber by including technology as a building block of the cage.

Since the 1960s, however, social movements have challenged many of the technical systems that control modern societies. These movements are oriented toward ends such as human rights and the health and well-being of humans and nature. Whether such ends can be shown to be rational by philosophers, they certainly present a very different picture from the pessimistic view of Weber and the Frankfurt School. Furthermore, they call into question the competence of technical systems, which is vastly overemphasized in technocratic ideology. There are reasons to doubt Weber's assumption that bureaucracy is the ideal form of instrumentally rational administration. This assumption underlies the technocratic dismissal of citizen interventions as irrational and inefficient.

One form of citizen intervention has attracted the interest of philosophers because it appears to correspond to influential theories about rationality. These theories challenge the view that ends are inherently nonrational. Scientific procedures yield one type of rationality, but there is another type based on free discussion. Debate over values and worldview also has a rational

character where the arguments are framed by procedures that are disinterested, inclusive, and respect the rights of all concerned.

Habermas's formulation of this notion as "communicative rationality" lies behind the current interest in deliberative democracy. There are attempts to apply something like the notion of deliberative democracy to technical controversies, for example, to show the democratic character of 'citizen juries' in technological decision making. These are small conferences of citizens and experts called to consider the wisdom of specific technical policies. Although they remain marginal, they have had an impact especially in Europe (Zhao et al. 2015; Fuller 2006, chapter VI). One famous such conference dissuaded the Norwegian state from authorizing genetically modified crops (Sclove 1997). Citizen juries illustrate the virtues of rational discourse in the determination of policy and suggest the possibility of public understanding of and intervention into the technical disciplines.

This is an interesting application of the concept of democratic rationality, but it represents only one of many forms of public intervention in the technical sphere. In its original formulations and in the writings of many commentators, this theory ignores the technification of politics in modern societies. This misses Weber's point, not to mention that of later critics, including Habermas's own early work on the public sphere. The problem is not so much that ends are inherently nonrational as that the technification of politics makes them so. In that case, the theory must address the real world of democracy in which social movements must contend with a system of mass media biased toward domination. The activities of democratic citizens must therefore include demonstrations, hacking, boycotts, lawsuits, and other means of impressing the public will on recalcitrant institutions. Most discussions of rationality in political theory follow Habermas in emphasizing the role of argument to the exclusion of the many other ways in which political debate is carried on.

A broader theory of the rationality of these various modes of intervention is needed to better evaluate the potential of technical citizenship. It would address the role of such things as agenda setting, acknowledgement of facts, testimony to the seriousness and priority of issues, and demands for participation. These are matters that are settled in the public sphere through action rather than argument, but they cannot be dismissed as nonrational without vitiating the concept of citizenship. I believe that a theory can be developed in which the rationality of such procedures is recognized, although this is not the place to attempt that task (Ingram 1987: 184)). Instead, I will consider ways in which technical citizenship can improve the instrumental rationality of modern societies.

Weber's still-influential views on bureaucracy constitute a second obstacle to a theory of democratic technical citizenship. Weber assumed uncritically that better calculation and control imperatively required bureaucratic administration. His model was the German bureaucracy and corporate management of his day. As a result, his theory of rationalization led to a pessimistic conclusion. Marcuse pointed out in an important article that Weber simply assumed his conclusion, that rational organization must divide conception from execution as it did in the institutions he took as models (Marcuse 1968). Modern sociology and business commentary have challenged Weber's conclusions from many angles, often arriving at the conclusion that successful management can be inclusive and participatory. Innovation, another important feature of modernity, certainly requires more freedom than a Prussian bureaucrat would normally have allowed, and this too is a widely held view among critics of bureaucracy.

We need a generalized rationalization theory that follows Weber in affirming the importance of calculation and control, but drops his insistence on bureaucracy as the only rational form of administration. Theories of democratic socialism and participatory capitalism assume some version of such a generalized rationalization theory, but the actual technical politics emerging today is far less ambitious than these utopian schemes. It is not systematic but takes the form of

democratic interventions, punctual initiatives from below tied to particular cases at particular times and places.

The concept of democratic intervention describes cases where the public becomes involved in conflicts over technology: for example, controversies over pollution or medical treatments (Callon et al. 2011). New regulations and standards often emerge from these controversies as public demands are assimilated by the existing system. Public participation in design is a second mode of intervention. This approach especially characterizes the computer industry where there is frequent consultation with users in the creation of new programs. I call a third mode of intervention the creative appropriation of technology, reinventing devices to meet new demands. The most impressive such case is, of course, the internet. The basic framework was supplied by the government but reworked by innovative users with technical skills. Their innovations include essentially all the communicative applications of the network. The fact that these innovations were widely adopted by the user community gives them a democratic character.

None of these interventions aim at abolishing bureaucracy, but they all bend the bars of the “iron cage” to allow in the effects of reflection on experience and creative initiative. It is reasonable to call these interventions *rationalizations* where they effectively improve technological efficiency relative to a socially accepted goal of some sort. The effect may not be visible from the standpoint of specific corporations or government agencies. They often pay a price to conform with new public demands. We hear their protests in the name of ‘efficiency’ all the time. But if the efficiency of the technological system is measured from the standpoint of society as a whole, then it is clear that interventions for such things as pollution control or improved opportunities to communicate do constitute technical progress.

Micro-political activism of this sort is the specific form of agency associated with technical citizenship. Micro-politics differs from such large-scale interventions as elections and revolutions that aim at state power. It may lack long-term organization, and it is often focused on a single issue and sometimes a single location. Nevertheless, the effects of micro-politics are widely felt. Democratic interventions alter the designs of particular artifacts and whole domains. After the effects of the interventions are translated into technical language and incorporated into the technical disciplines, the social forces behind the changes are forgotten. Today we are in the midst of such changes and can better see how society and technology interact. For example, public awareness of climate change has created pressures to transform energy production. Democratic interventions have been particularly significant in recent years in the defense of the internet against some of the most egregious forms of exploitation and invasion of privacy. This is a special and irreplaceable form of activism in a technological society. It limits the autonomy of experts and forces them to redesign the worlds they create to represent a wider range of interests.

Critical Theory of Technology

These observations call into question many old ideas about technology. Assuming that the argument so far establishes the reality of technical agency, we need a new theory to explain its conditions of possibility, freed from the influence of technocratic ideology. In what follows I am going to sketch my version of such a theory.

As noted earlier, we tend to think of scientific and mathematical ideas as independent of interests and personal preferences, and this assumption is often extended to technology as well. It is indeed difficult to track the influence of society on ideas in these fields, as opposed to law, politics, art, customs, and entertainment where social influences are obvious. This has given rise to the idea that technical rationality is ‘pure’ and value-neutral, a key premise of the technocratic view. But this concept of pure rationality is now contested in the field of science

and technology studies. Social history and sociology of technology routinely show the role of values in the design process. Where previously it was generally assumed that technologies were designed in response to purely technical considerations, more recent research demonstrates that many technical decisions are made in response to ideologies, visions of life, and interests.

Constructivist technology studies argues that what it calls “social actors” play an essential role in the design of technologies and technical systems. The notion of “actors” dereifies technical practice by restoring human decision in the technical sphere. Technology studies shows this concretely through research on particular cases of all sorts, from the history of bicycles to refrigerators, plastics, and vaccines. The old technological determinism that underpins the technocratic view is effectively refuted. Technology, it turns out, is ‘underdetermined’ by technical constraints. There is always room for alternative designs with different social consequences.

Technology is a hybrid of knowledge about nature, conserved in the technical disciplines, and the many concerns of nontechnical actors who intervene in design in a variety of ways. Several concepts have been introduced to signify the hybrid character of technology and the technical disciplines that create it. Wiebe Bijker proposes the concept of “technological frame,” a kind of paradigm or model that guides the actions and interactions of the many individuals and groups who cluster around the process of development. The similar notion of the “technological regime” has been developed by Arie Rip in a constructivist approach to technology assessment. I have suggested the term *technical code* to refer to the translation of social demands into technical specifications. Technical codes are incorporated into both designs and technical disciplines. Technical choices that depend on a social criterion of some sort bear a social content in technical form (Bijker 1995; Rip, Misa, and Schot 1995; Feenberg 2010).

In sum, there is no pure form of technology, designed in response to scientific knowledge and the pursuit of efficiency alone. Of course both science and efficiency are involved in design, but the outcome is also shaped by the constellation of social forces. Technologies are rational only within the framework laid out by the bias imposed on them by those forces. This is why the technocratic dismissal of social interventions as irrational intrusions is fundamentally wrong. There are always social interventions, for better or worse.

The technology of our contemporary world has a history. It is not the product of pure scientific knowledge but emerges from the industrial revolution of the nineteenth century. It was developed under the control of capitalist enterprise, which, uniquely among forms of property ownership, liberates the owner from responsibility to workers and communities. As a result, indifference to intolerable working conditions and destructive side effects characterizes the original technical codes of industry. This sets the stage for the labor struggles of the nineteenth century and the struggles over regulation in the twentieth century. The socialist movement has always regarded these struggles as preliminary to a general overthrow of the property system, but their structure indicates a more general feature of technological societies that is significant under any regime. The relation of lay and expert actors is implicated not only in the challenge to the private property, but also in every aspect of modern social life that is organized around technical disciplines and controls.

The valiative bias of technology reflects and evokes what I call *participant interests*. Technologies enroll individuals in networks. These networks associate the individuals in various roles, such as user, worker, or victim of its side effects. Each role implies certain interests and opens the individuals to a corresponding situated knowledge of the network. Where these interests are poorly served by the existing designs, tensions can arise, leading to demands for change. Because the individuals are on the inside, participating in the activities the technologies support, they sometimes have unusual insights and influence.²

The technical disciplines respond to these public interventions by gradually incorporating a broader range of considerations in their concept of the object and their practices. The increased

complexity of the disciplines corresponds to the real complexity of the world on which they operate. Values do not appear within the disciplines directly and immediately, but indirectly through designs that address side effects and opportunities identified in protest, hacking, and innovation. The results are improved technical disciplines and technologies as judged from both a technical and a normative standpoint.

The dynamic interaction of technical disciplines and public interventions is a consequence of the differentiations that make modernity possible. Until technology achieved a certain autonomy it was constrained by craft traditions and religious, ethical and aesthetic conceptions. Modernity unleashes technology by creating the conditions for the development of independent formal rational systems. Social rationality is thus a productive feature of modern societies despite the problems it brings in its wake. The rapid development of modern technology, especially since World War II, has such vast and threatening impacts that technology can no longer conserve the full degree of autonomy it acquired in modern times.

The ultimate reality test for technology is public acceptance since the public must deal not only with each particular technology in its ideal setting but all of them together in the chaotic world of daily life. Informal experiential knowledge is not simply inferior to technical rationality but encounters the world differently. Rather than abstracting toward a limited explanatory register consisting of causes and functions, experiential knowledge connects ideas by association. It operates with the lived experience of the phenomena, where values and meanings circulate freely. This may lead to confusion and error, but it is also open to the unforeseeable complexity of the world. That openness can be essential where the narrow scope of the values and connections built into technology leads to disaster or blocks progress.

Feedback from 'reality' as it is experienced by ordinary people under these conditions is thus not extraneous to technology but essential to its successful development. In a differentiated society, that feedback takes place through a sometimes conflictual circulation of information and products between technical experts and their technical disciplines and society at large. In sum, neither technical rationality nor everyday rationality are complete in themselves: they form halves of a fragmented whole. Anything that promotes the interaction of these divergent ways of understanding the world is progressive. Obstacles to communication between technical specialists and those affected by their activities should be removed as quickly and as effectively as possible. That is the new task of the democratic process that has been unfolding in the modern world since the eighteenth century in parallel with technical advance.

Notes

1 See also chapter 22, by Dana Belu, in this volume.

2 For examples of user agency, see Grimes and Feenberg (2012) and Hamilton and Feenberg (2012).

References

- Abbate, J. (1999). *Inventing the Internet*. Cambridge, MA: MIT Press.
- Bijker, W. (1995). *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press.
- Callon, M., Lascoumbes, P., and Barthe, Y. (2011). *Acting in an Uncertain World*, trans. G. Burchell. Cambridge, MA: MIT Press.
- Epstein, S. (1996). *Impure Science: AIDS, Activism, and the Politics of Science*. Berkeley, CA: University of California Press.
- Feenberg, A. (2010). *Between Reason and Experience: Essays in Technology and Modernity*. Cambridge, MA: MIT Press.

- Feenberg, A., and Freedman, J. (2001). *When Poetry Ruled the Streets: The May Events of 1968*. Albany: SUNY Press.
- Fuller, S. (2006). *The Philosophy of Science and Technology Studies*. New York: Routledge.
- Grimes, S., and Feenberg, A. (2012). Rationalizing Play: A Critical Theory of Digital Gaming. In A. Feenberg and N. Friesen (eds.) *(Re)Inventing the Internet*. Rotterdam: Sense Publishers, 21–41.
- Hamilton, E., and Feenberg, A. (2012). Alternative Rationalisations and Ambivalent Futures: A Critical History of Online Education. In A. Feenberg and N. Friesen (eds.) *(Re)Inventing the Internet*. Rotterdam: Sense Publishers, 43–70.
- Ingram, D. (1987). *Habermas and the Dialectic of Reason*. New Haven, CT: Yale University Press.
- Lievrouw, L. (2011). *Alternative and Activist Media*. Cambridge: Polity Press.
- Marcuse, H. (1968). Industrialization and Capitalism in the Work of Max Weber. In J. Shapiro (trans.) *Negations: Essays in Critical Theory*. Boston, MA: Beacon Press.
- Michaels, P. A. (2014). *Lamaze: An International History*. Oxford: Oxford University Press.
- Pinch, T., and Bijker, W. (1987). The Social Construction of Facts and Artefacts. In W. Bijker, T. Hughes, and T. Pinch (eds.) *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press, 17–50.
- Rip, A., Misa, T., and Schot, J. (eds.) (1995) *Managing Technology in Society: The Approach of Constructive Technology Assessment*. London/New York: Pinter.
- Sclove, R. (1997). *Democratizing Science and Technology*, lecture given as part of the University of Massachusetts Lowell Center for Competitive Enterprise Lecture Series, December 5.
- Wagner, P. (1998). Sociological Reflections: The Technology Question During the First Crisis of Modernity. In M. Hård and A. Jamison (eds.) *The Intellectual Appropriation of Technology*. Cambridge, MA: MIT Press, 225–252.
- Weber, M. (1958). *The Protestant Ethic and the Spirit of Capitalism*, trans. T. Parsons. New York: Scribners.
- Winner, L. (1992). Citizen Virtues in a Technological Order. *Inquiry*, 35(3/4), 341–361.
- Zhao, Y., Fautz, C., Hennen, L., Srivivas, K. R., and Li, Q. (2015). Public Engagement in the Governance of Science and Technology. In M. Ladikas, S. Chaturvedi, Y. Zhao, and D. Stemerding (eds.) *Science and Technology Governance and Ethics: A Global Perspective from Europe, India and China*. New York, Dordrecht, London: Springer Open Access.

Animals in Philosophy of Technology

Ashley Shew

Introduction

Technology is typically discussed as a human-only phenomena. In discussions of technology, historians, sociologists, philosophers, and other storytellers often take as their starting points either the dawn of human civilization, the industrial revolution, or the computer age. They set chronological limits on the domain of inquiry about technology, trying to make sense of the promises and perils of today's technological milieu. Oftentimes, philosophers of technology focus on modern engineering practice and proliferation, analyzing technology through a contemporary engineering frame.

Definitions of technology reflect this frame. Historians Mevlin Kranzberg and Carroll Pursell define technology as “man’s effort to cope with his physical environment . . . and his attempts to subdue or control that environment by means of his imagination and ingenuity” (1967). Forbes writes that “[t]echnology is as old as man himself” (1964). Leinhard explains that, “[t]echnology has driven our brains . . . [and] has, in turn, expanded our minds and fed itself” (2003). Joseph C. Pitt defines it as “humanity at work” (2000), while Thomas P. Hughes sees it as “a creative process involving human ingenuity” (2004). But, Keekok Lee writes that “technology—in the general sense of the manipulation of nature to suit human purposes—is not peculiar to modernity” (2009).

All of these definitions, and many others, associate technology with humanity. “Technology made the man, and man made the technologies,” in the words of Kranzberg and Pursell (1967). In this way, philosophers of technology have neglected animal research and creation. The last thirty years of animal research suggest that the sorts of hallmarks we closely relate to human invention can also be found within the wider animal kingdom: researchers have found evidence for great intelligence, design, innovation, material cultures, causal reasoning, and the use of metatools among other species.

The past thirty years of animal research has shown tool use and construction are much more widespread and complex than has been appreciated by philosophers. New insights in animal science provide better grounding for philosophers of technology to take up questions about the nature of tools and technology, the role tools serve in the lifeworlds of their users, the roles of animals in our thinking about human beings, and more. Animal studies provide a fresh perspective that might bring balance to what is right now a very engineering-focused discipline, and an appreciation of animal studies within the discipline might lead to wider audiences and concerns.

While studies of engineering remain a staple of interest in philosophy of technology, animal studies deserves more investigation, for animals make and use things too; they are also, in the age of big biotechnology, the first groups of participants in our trials. Our environmental disasters that come from our technological hegemony impact animal survival, diversity, and life. Moral philosophy has already taken up the cause of animals, with the ethics of animal consumption,

their use by science, and conditions of their lives scrutinized (Beckoff 1972; Singer 1972). Philosophers of technology provide a different approach by looking at how tool use varies, the conditions under which creatures construct objects, and how techniques evolve, as well as more traditional questions about the relationship between humans and nonhumans, the defining features of technology (vs. tool use), and the use of animals as technologies. This chapter argues that we need to consider animal cases and suggests how that might be accomplished.

Many philosophers of technology take as their starting point the industrial revolution or the invention of the steam engine to mark off modern technology from premodern (Borgmann 2009; Heidegger 2010). I would argue that there is no easy dividing line, and that, if we are to investigate the meaning and importance of technology, writ large, looking at animal cases ought to be focal in any philosophy of technology. Not just a contrast class, animal studies provide insights into our embodiedness, our evolution as tool users, and our relationships to other creatures, as well as notions like creativity, invention, technological knowledge, and innovation. Neglecting the animality of our human enterprises no longer seems reasonable, given the state of animal research and given the importance of technology in our lives. To ask ontological, phenomenological, epistemological, and ethical questions about technology requires consideration of the meaning and evolution of tool use to technology.

Themes and Motivations

Breaking free of the cultural underpinnings to our understandings of animals requires much work. Oftentimes, through biblical and other cultural mythologies, nonhuman animals get situated as inferior to humans, as in need of good stewards in humans, and as lesser in important human capacities. In Plato's tripartite theory of the soul, it is only humans who possess all three parts: logical, spirited, and appetitive; Aristotle also sets humans apart from nonhuman animals, who supposedly lack the capacity for reason that humans have. This idea that humans (and particularly male humans) are rational, while animals (and sometimes women) are not rational dominates much of our history. In biblical stories, we witness Noah saving all the animal species that exist by loading them, two-by-two, onto his ark (Genesis 7:9); stewardship and care becomes the frame through which to view nonhumans. Through this dominant cultural lens, the evaluation of animal intelligence and capacity was understood such that many of the things animals make and do—from nest building to hunting—get reduced to mere instinct instead of indicating learning, intelligence, or creativity. Add to this the paradigm of behaviorism directing all understanding of mind by the 1950s, where input-output machines become the metaphor for animal life, and we've set ourselves up for misunderstanding, preconceived notions, and bad science and philosophy.

Donna Haraway highlights that bad science in her *Primate Visions* (1989), and later introduces the notion of the cyborg to slip out of these narratives, to liberate women (and animals) from these hegemonic frames (Haraway 1991). Peter Singer and Mark Beckoff have worked to ameliorate the horrible idea, which comes from our cultural mythologies, that animals either don't suffer or that their suffering doesn't matter (Singer 1972; Beckoff 1972). In moral philosophy, we see the most robust treatment of animals within the philosophical literature. Peter Singer's *Animal Liberation* (1972) was a utilitarian-ethics-based call to end animal suffering. By counting animal suffering as real and important to our moral evaluations, he elevates animals for consideration within moral theory. Mark Beckoff has stressed the same: animal suffering is of significant moral consequence. Working from a deontological frame, Onora O'Neill has integrated animal concerns as well (O'Neill 1997). The idea that animals must be factored in ethical considerations has had an impact on bioethics, the treatment of laboratory animals for research, and our continued use of animals for meat and subservience to humans. They have

effectively argued that animals matter in ethics, though they might disagree on the specific details of those theories.

The emergence of philosophy of biology as an important branch of science for philosophical consideration means that there is now ample work on evolutionary mechanisms, language, and methodology that involves animal subjects (Dawkins 1999; Millikan 1984; Burian 1978). Despite increasing sophistication in animal theory and differentiation between species and cognitive properties, philosophy of technology has not used animal cases in a rich way. Those scholars working on AI, philosophy of mind, and biotechnology sometimes analogize to animal cognition or construction, using animal cases as proof that something is possible or inevitable when it comes to engineered intelligence (Dennett 1995a; Dennett 1995b; Searle 1998). Thomas Nagel, in “What is it like to be a bat?” explores how we might explore animal thoughts and the difficulties in that process (Nagel 1974).

In terms of ethics for emerging technology and modern life, Paul Thompson employs animal cases to discuss the difficult decisions that lie ahead in the treatment of farmed animals, given the rise of genetic engineering (Thompson 2013). Diane Michelfelder has considered nonhuman animals in the context of cities (Michelfelder 2003).

Philosophers of technology have sometimes made reference to animal cases, but not in a rich way outside of their use in ethics, and otherwise still often taking animals as the sort of input-output machines seen in behavioristic theory. Insisting on the difference between human and animal artifacts, Davis Baird writes that spider webs “are well adapted to catch flies . . . there is no connection established between this approach to catching spider food and other possible and actual approaches” (2004: 141). By using this case, he dismisses a whole class of potential knowledge-bearing constructions, the subject of Baird’s work. The use of the example of the spider serves to underappreciate those objects constructed by animals that do demonstrate sophistication, like the use of a set of tools and shaping techniques by New Caledonian crows (Hunt and Gray 2004) or the material culture of different groups of chimpanzees (Sanz and Morgan 2007).

Similarly, Joseph C. Pitt defines technology as “humanity at work” (2000), but leaves room that the definition might some day include “the activity of beavers or alien” if we develop a good sense of what constitutes “purposeful activity for non-humans” (2000: 11). I think new animal research indicates that we have some sense of work on the part of other species as being more than what mechanistic or behavioristic models would lead us to believe. Important work has been done by Mieke Boon about how we care for other creatures (2005), and this is a start in considering how we use animals in technological contexts. What I suggest here is that we need to engage with animal studies on several levels, including this one.

Objections

Animal cases belong in any study of epistemology or intentionality with regard to the use of technologies. The things nonhuman animals construct are worth our consideration, especially if we hope to develop rich understandings of how and why our technologies take the forms they do today. More than just satisfying needs or desires, the ways in which different groups use tools differently, pass on tool-making practices, or fail to use tools, provides for enriching discourse about the trade-offs in technology adoption and use, as well as an understanding of function and form suited to particular tasks. I think here of studies of the tool use of crows and rooks (Hunt and Gray 2002; Weir et al. 2002; Taylor et al. 2007; Bird and Emery 2009), of dolphin techniques and specialization (Krutzen et al. 2005; Sargeant, Mann, Berggren, and Krutzen 2005; Gazda et al. 2005; Pryor and Norris 1991; Janik et al. 2006), and of primate behavior and tool use (Goodall 1986; Sanz and Morgan 2007; Toth and Schick 1993;

Inoue-Nakamura and Matasuzawa 1997; Osvath and Osvath 2008; Boesch and Boesch 1989; Visalberghi et al. 1995). These cases provide rich fodder for thinking about how techniques are communicated, how practices with tool use are culturally variant, and how knowledge about objects is made and transmitted. In exploring cases of animal tool use and construction, I've encountered many objections to the inclusion of animal cases in discussions of epistemology of technology (although I suspect these objections would also be targeted in using animal cases in other areas). Here are the four main objections to their integration and replies to each.

Objection 1: "You're Just Anthropomorphizing"

Anthromorphization is inappropriately mapping human characteristics onto animals. The force behind this objection is that, in fact, in studies of things like cognition, tool making, and reason, one does indeed map human concepts onto animal behavior. These are concepts we consider important to humanity, to how we define ourselves as a species, but so much of this is bound up with cultural narratives about what humans do and think vs. what animals do and think. However, not mapping these concepts on, when appropriate, is also problematic. Ted Kerasote writes:

[a]nthropomorphization is often maligned for ascribing human characteristics to animals who can't possibly know what we know. And there is some truth to this. I doubt [my dog] thought of the Big Bang when he gazed at the starry heaven. But the reverse—not ascribing volition to creatures who repeatedly display it—is also inaccurate. It leads to what poor translation always does: misunderstanding between cultures.

(Kerasote 2008: 112)

To not project our concepts onto some of the ways in which animals behave presents a problem—and, by not using some of our concepts, we actually impoverish our own worldviews. The idea that animals couldn't suffer or feel pain in the way humans do helped justify neglect and mistreatment of animals; while moral philosophy has taken up the case of animal suffering (they suffer as we do), philosophers would also do well to take up the idea that many animals demonstrate remarkable intelligence, problem solving, and cognitive function. While it may not always be 'like us,' and accepting that proving these things are difficult to demonstrate, scholars and researchers should project and maintain that some of the things nonhuman animals think and feel are parallel to the things that humans do.

Objection 2: "We Need to Understand Humans First"

While some philosophers are friendly to accepting animal cases as important and related to what humans do, they still push for looking solely at human invention in our technological studies. Similar for justifications that pare philosophy of technology down to just looking at engineering cases, these scholars argue that we know for sure that humans use technologies and tools in a sophisticated way, while we cannot be sure of what is going on with animals. (The engineering analogue is: we can't be sure that every material product fits into the category of 'technology,' but we know the products of engineering do.) The thinking goes that, to investigate the nature of technology, we need to set a domain of discourse, and animal products are unclear in this regard.

However, if we truly wish to investigate the nature of technology, whatever we mean by that, looking at the challenging and border cases would actually help enrich a study and give clear guides. In my own scholarship, I've discussed how beavers building dams provide less clear

intentional behavior than, say, bowerbirds do in constructing their nests or New Caledonian crows do in the making of their tools (Shew 2007; Shew 2008). By taking into account a greater spectrum of activity, we can appreciate nuance and provide more clarity in terms of the origins of technology and tool use as activities. I would argue that any epistemology or ontology of technology needs reference to the evolutionary and social processes by which technologies are created, maintained, and distributed.

Objection 3: “We Can’t Possibly Understand Animal Cognition”

We can’t understand animal cognition, but it’s fair to infer certain processes, like the ability to plan and shape and think abstractedly, when animal studies are carefully done and defined. Appreciating that other animals do think and make and communicate is as important to recognizing that other humans do too. Just because those cognitive powers manifest differently does not mean that they aren’t there. For too long, humans have taken animal cases as trivial and have taken animal minds as insufficient and inferior. To continue assuming that we should bracket off human behavior as special while not investigating the wider spectrum of behaviors and minds seems absurd. Surely, nonhumans are not directed or conditioned for the things humans are, but that doesn’t make their work or lives outside the sphere of technological study.

Mike Hansell provides wonderful accounts of animal construction. One of his chapters illustrates something so important to remember: “Tools Aren’t Always Useful” (Hansell 2007). It’s funny that we take tool use and construction to be of utmost important and the clearest indicator of intelligence when, for most animal species, carrying an extra something around with them actually confers a disadvantage to survival. There are plenty of smart species out there that have developed no tool-using behaviors—and that indicates more about the ecological niche that they inhabit than it does the function of their minds. There have been some amazing studies comparing the tool-use of rooks to New Caledonian crows. New Caledonian crows use a dizzying array of tools in the wild, and tests in captivity have shown that rooks, who make no tools in the wild, are just as adept at problem solving and making tools when put through the same tests (Bird and Emery 2009). While tool use won’t always indicate animal cognition, the material products and capacities for manipulation of objects can provide a good study in how (and in what niches) tool use can arise; this can lend itself to a better understanding of the evolutionary mechanisms at work in the rise of *homo sapiens* (man the knower) as *homo faber* (man the maker).

We assume that other humans think like we do; we make inferences about other human cognition all the time. Assuming the same about some animals, especially given the sophistication of research on animal intelligence, seems to be less problematic than assuming that our lack of understanding about cognition in general should keep us from using animal cases, especially when our interest involves material production (something you can hold without probing into another’s mind).

Objection 4: “Animals Have No Concept of Technology, Therefore We Shouldn’t Project It Onto Their Activities”

I have three replies to this objection. We humans have made the classification system and developed the concepts, but that doesn’t mean they cannot be applied elsewhere. First, we do this all the time with technology. We classify plenty of things from the recent and distant past as ‘technology’—the wheel, swords, trebuchets, and buggy whips—when the people who once used and made those things did not use the word or contemporary category of ‘technology’ in the way we do now. We don’t generally consider this a problem; we talk about the history of

technology, referring to the whole of made objects, without contradiction or irrelevance. (The problems generally arise when we assume everyone thought like us or that our categories are real and unmovable, instead of convenient ideas for which to categorize things.)

Second, being able to understand a category isn't a requirement for doing something within that category. Marginal cases bear this out: a baby can roll over and scoot and crawl, for instance, without knowing the category or the name of those actions. Knowing what you are doing is not a necessary condition for doing it.

Third, there's also a precedent, not only in discussions of technology, but also in other academic work for reterming a group of ideas to better make sense of how people did things in the past. Jessica Riskin, in *Science & the Age of Sensibility* (2002), uses the concept of "sentimental empiricism" to trace an intellectual history of the French enlightenment. Though this is not a term adopted by the thinkers she discusses, the idea she introduces helps to explain the literature she tracks, demonstrating how modern science developed from these once-seemingly conflicting concepts of reason and sensibility. The concepts we use are supposed to help us explain things, whether or not the actions or constructions we hope to explain use those same categories or not. Crows could still shape twigs, and spiders could still spin webs, and dolphins could still use signature vocal tunes without those concepts existing in human minds.¹ Importing our concepts onto their activities is in no way a problem as long as we don't also assume that they share all of our categories.

Agenda

By now you are convinced that animal cases *can* be incorporated, but why *should* you use them? Scholarly conversations about technology take many forms, from reflection about Big Technology to intellectual property and nanotechnology. Animal cases won't always make sense. However, in the sort of core issues that pop up in topics in the subject, animal cases can be integrated. Aside from bioethics (which has already integrated concerns about animals), core topics include the topics that the discipline of philosophy of technology grew up around: the transmission and nature of technological knowledge, the role of skill and tacit knowledge, the differences between science and technology (science being a very particular cultural practice), the nature of humanity in the face of technology (where animal cases might provide relief), and the defining features and nature of technology itself. Some of this has been taken up in AI literature (in making sense of intelligence), but there is much left to do, especially in amending fundamental concepts and disputes in philosophy of technology to make use of these cases.

Animal research has demonstrated that humans are not the only makers-of-things; it has also shown tool use (and systems of tool use and technique) to be situated in environmental niches and cultural contexts. Continuing to ignore these cases and solely turn to engineering for guidance about the nature, transmission, and importance of technology in our lives no longer remains feasible if we wish to engage in broader reflection about technology.

Additionally, doing the kind of scholarship I suggest—with rich incorporation and appreciation of animal cases—speaks to and incorporates larger communities of researchers who wish to engage on broad questions about intelligence, intentionality, creativity, innovation, material culture, and social contexts. Philosophy of technology is uniquely situated to take up questions about these things, with a rich set of cases from engineering already at hand; by contrasting those cases with a broader set, philosophy of technology's relevance becomes clearer. Philosophy of technology is, arguably, the most important area in philosophy, whether or not philosophers in general recognize that. Reflections about technologies—indeed, about how our lives are framed and understood—help do exactly what Wilfred Sellars suggested was the definition

of philosophy: “The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term” (1963). Making sense of how things “hang together” involves looking at more things, and philosophy of technology already has the frames for investigation ready to fan out onto animal cases.

Note

- 1 There’s actually evidence that dolphins can understand abstract categories. Some search and rescue dolphins in captivity are able to go searching underwater for man-made objects with a particular size range with instructions that point to this vague category of stuff (Pryor and Norris 1991); there’s also evidence that dolphins have signature vocal tunes and something like naming (Janik et al. 2000 Janik et al. 2006). It’s not clear that dolphins can’t have the concepts we are using in their practice.

References

- Baird, D. (2004). *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley, CA: University of California Press.
- Bekoff, M. (1972). The Development of Social Interaction, Play, and Metacommunication in Mammals: An Ethological Perspective. *Quarterly Review of Biology*, 47(4), 412–434.
- Bird, C. D., and Emery, N. J. (2009). Insightful Problem Solving and Creative Tool Modification by Captive Nontool-Using Rooks. *Proceedings of the National Academy of Sciences*, 106(25), 10370–10375.
- Boesch, C., and Boesch, H. (1989). Hunting Behavior of Wild Chimpanzees in the Tai National Park. *American Journal of Physical Anthropology*, 78(4), 547–573.
- Boon, M. (2005). Comments on Thompson: Research Ethics for Animal Biotechnology. *Frontis*, 5, 121–125.
- Borgmann, A. (2009). *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago: University of Chicago Press.
- Burian, R. M. (1978). A Methodological Critique of Sociobiology. In Arthur L. Caplan (ed.) *The Sociobiology Debate*. New York: Harper & Row, 376–395.
- Dawkins, R. (1999). *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press.
- Dennett, D. C. (1995a). Animal Consciousness: What Matters and Why. *Social Research*, 62(3), 691–710.
- Dennett, D. C. (1995b). Do Animals Have Beliefs? In H. L. Roitblat and Jean-Arcady Meyer (eds.) *Comparative Approaches to Cognitive Science*. Cambridge, MA: MIT Press, 111.
- Forbes, R. J. (1964). *Studies in Ancient Technology*. 5, Brill Archive.
- Gazda, S. K., Connor, R. C., Edgar, R. K., and Cox, F. (2005). A Division of Labour With Role Specialization in Group—Hunting Bottlenose Dolphins (*Tursiops truncatus*) off Cedar Key, Florida. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1559), 135–140.
- Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, MA: Belknap Press of Harvard University Press.
- Hansell, M. (2007). *Built by Animals: The Natural History of Animal Architecture*. Oxford: Oxford University Press.
- Haraway, D. J. (1989). *Primate Visions: Gender, Race, and Nature in the World of Modern Science*. New York: Psychology Press.
- Haraway, D., and Teubner, U. (1991). *Simians, Cyborgs, and Women*. na.
- Heidegger, M. (2010). The Question Concerning Technology. In Craig Hanks (ed.) *Technology and Values: Essential Readings*, Chichester, UK and Malden, MA: Wiley Blackwell, 99–113.
- Hughes, T. P. (2004). *Human-Built World: How to Think About Technology and Culture*. Chicago: University of Chicago Press.
- Hunt, G. R., and Gray, R. D. (2002). Species-Wide Manufacture of Stick-Type Tools by New Caledonian Crows. *Emu*, 102(4), 349–353.

- Hunt, G. R. and Gray, R. D. (2004). The Crafting of Hook Tools by Wild New Caledonian Crows. *Proceedings of the Royal Society*, London B (Suppl.) 271, S88–S90.
- Inoue-Nakamura, N., and Matsuzawa, T. (1997). Development of Stone Tool Use by Wild Chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 111(2), 159.
- Janic, V. (2000). Whistle Matching in Wild Bottlenose Dolphins. *Science*, 289, 1355–1357.
- Janik, V. M., Sayigh, L. S. and Wells, R. (2006). Signature Whistle Shape Conveys Identity Information to Bottlenose Dolphins. *Proceedings of the National Academy of Sciences*, 103(21), 8293–8297.
- Kerasote, T. (2008). *Merle's Door*. New York: Houghton Mifflin Harcourt.
- Kranzberg, M., and Pursell Jr., C. W. (1967). The Importance of Technology in Human Affairs. *Technology in Western Civilization*, 1, 10.
- Krützen, M., Mann, J., Heithaus, M. R., Connor, R. C., Bejder, L., and Sherwin, W. B. (2005). Cultural Transmission of Tool Use in Bottlenose Dolphins. *Proceedings of the National Academy of Sciences of the United States of America*, 102(25), 8939–8943.
- Lee, K. (2009). Homo Faber: The Unity of the History and Philosophy of Technology. In an Kyrre Berg Olsen, Evan Selinger, and Søren Riis (eds.) *New Waves in Philosophy of Technology*. London: Palgrave Macmillan UK, 13–19.
- Lienhard, J. H. (2003). *The Engines of Our Ingenuity: An Engineer Looks at Technology and Culture*. Oxford: Oxford University Press.
- Michelfelder, D. P. (2003). Valuing Wildlife Populations in Urban Environments. *Journal of Social Philosophy*, 34(1), 79–90.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.
- Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435–450.
- O'Neill, O. (1997). Environmental Values, Anthropocentrism and Speciesism. *Environmental Values*, 6(2), 127–142.
- Osvath, M., and Osvath, H. (2008). Chimpanzee (*Pan troglodytes*) and Orangutan (*Pongo abelii*) Forethought: Self-Control and Pre-Experience in the Face of Future Tool Use. *Animal Cognition*, 11(4), 661–674.
- Pitt, J. C. (2000). *Thinking About Technology: Foundations of the Philosophy of Technology*. New York: Seven Bridges Press.
- Pryor, K., and Norris, K. S. (1991). *Dolphin Societies: Discoveries and Puzzles*. Berkeley, CA: University of California Press.
- Riskin, J. (2002). *Science in the Age of Sensibility: The Sentimental Empiricists of the French Enlightenment*. Chicago: University of Chicago Press.
- Sanz, C. M., and Morgan, D. B. (2007). Chimpanzee Tool Technology in the Goulougo Triangle, Republic of Congo. *Journal of Human Evolution*, 52(4), 420–433.
- Sargeant, B.L, Mann, J., Berggren, P, and Krutzen, M. (2005). Specialization and Development of Beach Hunting, a Rare Foraging Behavior, by Wild Bottlenose Dolphins (*Tursiops* sp.). *Canadian Journal of Zoology* 83, 1400–1410.
- Searle, J. R. (1998). How to Study Consciousness Scientifically. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1377), 1935–1942.
- Sellars, W. (1963). Philosophy and the Scientific Image of Man. *Science, Perception and Reality*, 2, 35–78.
- Shew, A. (2007). *Beaver Dams, Spider Webs, and the Sticky Wicket*. Blacksburg: Virginia Polytechnic Institute and State University.
- Shew, A. (2008). *What Dolphins Want: Animal Intentionality and Tool-Use*. Blacksburg: Virginia Polytechnic Institute and State University.
- Singer, P. (1972). Moral Experts. *Analysis*, 32(4), 115–117.
- Taylor, A. H., Hunt, G. R., Holzhaider, J. C., and Gray, R. D. (2007). Spontaneous Metatool Use by New Caledonian Crows. *Current Biology*, 17(17), 1504–1507.
- Thompson, P. B. (2013). The Opposite of Human Enhancement: Nanotechnology and the Blind Chicken Problem. In Sean A. Hays, Jason Scott Robert, Clark A. Miller, and Ira Bennett (eds.) *Nanotechnology, the Brain, and the Future*. Dordrecht: Springer, 247–263.

- Toth, N., and Schick, K. (1993). *Making Silent Stones Speak: Human Evolution and the Dawn of Technology*. New York: Touchstone.
- Visalberghi, E., Frigaszy, D. M., and Savage-Rumbaugh, S. (1995). Performance in a Tool-Using Task by Common Chimpanzees (*Pan troglodytes*), Bonobos (*Pan paniscus*), an Orangutan (*Pongo pygmaeus*), and Capuchin Monkeys (*Cebus apella*). *Journal of Comparative Psychology*, 109(1), 52.
- Weir, A. A., Chappell, J., and Kacelnik, A. (2002). Shaping of Hooks in New Caledonian Crows. *Science*, 297(5583), 981–981.

Virtual Space and Property



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Obscurity and Privacy

Evan Selinger and Woodrow Hartzog

Introduction

‘Obscurity’ is a distinctive concept in the privacy literature that has recently been gaining attention due to increasing frustration with the theoretical and practical limits of traditional privacy theory.¹ Obscurity identifies some of the fundamental ways information can be obtained or kept out of reach, correctly interpreted or misunderstood. Appeals to obscurity can generate *explanatory power*, clarifying how advances in the sciences of data collection and analysis, innovation in domains related to ICT, and changes to social norms can alter the privacy landscape and give rise to three core problems: (1) new breaches of etiquette, (2) new privacy interests, and (3) new privacy harms.

If citizens truly value opportunities to communicate over public channels and in public places while maintaining the protections obscurity can afford, they will need to marshal different types of resources than they currently possess. Contributions will need to come from humanities and social science researchers (especially philosophers), legal scholars and practitioners, technologists, designers, educators, and citizens themselves. Without multidisciplinary, multistakeholder, and multi-institutional commitment, a disconcerting outcome is likely to arise. Critics will continue to proclaim privacy dead and, perhaps justifiably, insist obscurity was eviscerated in the digital age. Fortunately, prematurely eulogizing privacy is a time-honored tradition that dates back to the advent of the penny press and the first Kodak handheld camera, but so far remains more a matter of pessimistic prognostics than accurate assessment of how society is structured.

Obscurity Fundamentals

In everyday conversations, the word *obscurity* designates unnoticed people and inconspicuous things. A typical newspaper story thus begins: “The first time I talked to Anita Hill, she was teaching commercial law in Oklahoma, living in obscurity in the state where she grew up on an isolated farm with 12 siblings” (Phelps 2014). In privacy scholarship, however, obscurity has a related, but different meaning (Hartzog and Selinger 2013a, b; Hartzog and Stutzman 2013a, b). Obscurity is the idea that information is safe—at least to some degree—when it is hard to obtain or understand.

When information is hard to come by, the only people who will seize upon it are those with sufficient motivation to expend the necessary effort and resources. Given the powerful influence of inertia on human behavior (Thaler and Sunstein 2009), the vast amount of information competing for our eyes and ears in today’s “hypertrophic attention economy” (Ricks 2013), and the prevalence of consumer devices designed to disburden us of hard work (Borgmann 1987), we should not underestimate how much of a deterrent effort can be.

When information is hard to understand, the only people who will grasp it are those with sufficient motivation to push past the layer of opacity protecting it. Sense-making processes of interpretation are required to understand what is communicated and, if applicable, whom the communications concerns. If the hermeneutic challenge is too steep, the person attempting to decipher the content can come to faulty conclusions, or grow frustrated and give up the detective work. In the latter case, effort becomes a deterrent, just like in instances where information is not readily available.

For example, you might overhear one person telling another at a café that “Immanuel is in therapy.” The use of a verb in the present tense suggests the person who is being discussed is alive, and not the long deceased Kant. But without further information, you will not know who, specifically, is the one seeing a psychologist, psychiatrist, social worker, or spiritual advisor. And even if that answer comes to be said aloud at a high level of volume, you still might not be able to decipher it. Perhaps the information is conveyed obliquely through an insider reference that means nothing to you, such as “our west coast friend.” At that point, you might conclude it is no longer worth it to continue eavesdropping.

The difficulty of identifying a person, notwithstanding the possession of relevant and intelligible information, is central to debates over obscurity-corrosive innovations such as facial recognition technology. While those around us can understand much of what we communicate in public, there can be a world of difference between others knowing what we say and knowing who we are when we say it.

In the digital age, obscurity-related issues frequently arise on the internet. There are several ways to make online communication more obscure: sharing ideas on platforms that are invisible to search engines; using privacy settings and other access controls; withholding your real name and speaking anonymously or identifying yourself with a pseudonym; disclosing information in coded ways that only a limited audience will grasp; or transmitting content that is encrypted or temporarily accessible through an ephemeral conduit, like Snapchat, the photo messaging application that can delete information within seconds after the recipient views it.

Given the prevalence of algorithmic surveillance today, obscurity practices go beyond making it difficult for other people to know what we’re saying or that we’re the ones saying it. They also include using strategies for conversing online “without tipping . . . intentions to the algorithm,” effectively “communicating without being computed” (Madrigal 2014). In the era of ‘big data,’ individuals are interested in remaining obscure from peers, strangers, and even algorithms. Some of the options to produce obscurity include: referring to folks without tagging them; referring to people and deliberately misspelling their names or alluding to them through contextual clues; sending screenshots of a story instead of directly linking to it; and, hatelinking, which introduces noise into a system by making it seem that you approve of a story, rather than denounce it (Madrigal 2014).

Because activities that promote obscurity can limit who monitors our disclosures without being subject to explicit promises of confidentiality, the tendency to classify information in binary terms as either ‘public’ or ‘private’ is inadequate. It lacks the nuance needed to describe a range of empirically observable communicative practices that exist along a continuum. For example, YouTube allows users to make published content public, private, or simply unlisted, which while not entirely ‘private,’ is less likely to be found. At one end of the *continuum of obscurity* lies information we’re shouting from the rooftops and want to be made absolutely transparent. At the other end of the continuum lies information we keep to ourselves and want to be kept absolutely secret. There is lots of middle ground in between these extremes. On many occasions, we try to communicate with one group or another in public and semi-public settings and hope the information does not travel past these groups aptly designated by the only seemingly oxymoronic phrase “private publics” (Jarvis 2011: 20).

By acknowledging the nuances of the obscurity continuum, it becomes possible to appreciate that many contemporary privacy debates are probably better understood if reclassified as concern over losing obscurity. The obscurity of public records and other legally available information is at issue in recent disputes over publishing mug shots, homeowner defaults, and a map of where gun owners live that was gleaned from public records. Likewise, claims for ‘privacy in public,’ like those that occur in discussion over license-plate readers, GPS trackers, facial recognition technologies, and emergent smart wearable devices like Google Glass, are often pleas for greater obscurity that get either miscommunicated or misinterpreted as insistence that one’s public interactions should remain secret. Even key controversies over social media—such as when Facebook rolled out Graph search and lifted its restrictions on public posts by teenagers—revolve around eroded obscurity, but erroneously get framed as matters concerning judicious use of privacy settings.

Ultimately, claims about how obscurity is gained or lost are descriptive propositions. On the one hand, obscurity theory does have something to say about the underlying causes that give rise to pro-obscurity attitudes: (1) there is evidence in evolutionary biology—highlighted by Robin Dunbar’s “social brain hypothesis”—that suggests natural limits constrain how much we can know about others without our memory systems becoming overburdened; and, (2) Erving Goffman’s sociology of everyday interaction suggests individuals regularly employ strategies that enhance obscurity in order to protect themselves and advance their goals (Hartzog and Stutzman 2013a: 6–8; Dunbar 1998; Dunbar 1993; Dunbar and Spoors 1995; Goffman 1959).

On the other hand, it is a normative question whether gains and losses of obscurity should be considered desirable or not. Consequently, addressing the prescriptive dimensions of obscurity requires looking to concepts developed in sources like moral and political philosophy for guidance. Consider an illustrative case that came up in an etiquette advice column (Vernon 2014; Hartzog and Selinger 2014).

A well-intentioned grandmother took screenshots of her grandkids’ freely accessible Instagram photos and proudly uploaded them to Facebook for all of her ‘friends’ on that social network site to see. This act hurt the grandkids’ feelings, and the grandmother could not grasp why they accused her of committing a faux pas. After all, she reasoned, the younger generation did not set their accounts to private, and all she did was lovingly pass along publicly available information.

It would appear that the grandkids were not upset because of a perceived loss of some traditional notion of privacy. Instead, they were most likely hurt because grandma eroded the obscurity of their photos and made it too easy for unwanted parties, likely family members, to find them on Facebook. Even though the pictures were publicly available on Instagram, presumably much of the family did not use that service—a contingent but important social reality, one that effectively kept the images hidden in plain view.

To go further and assess whether the grandmother’s behavior should be deemed a privacy violation, one would need to go beyond empirical fact finding (e.g., who uses what services) and subjective experiences (e.g., who feels displeased), and appeal to explicitly normative arguments. To get a sense of what this means, let us stipulate that the grandkids were dismayed because they believed the grandmother should have first consulted them before making any decisions that would influence who can easily access to their personal information. Because privacy is an essentially contested concept—or, as Daniel Solove puts it, a concept in “disarray”—different normative theories provide arguments that potentially could be used to justify this belief (2008: 1).

For example, the conviction that the grandkids have that their privacy had been violated could be justified by appealing to theorists who defend the paradigm that construes privacy as “control over personal information” (Solove 2008: 24–29). Alan Westin famously insists: “Privacy

is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others” (Westin 1967: 7). The belief also could be justified by appealing to theorists who defend the paradigm that defines privacy as “intimacy” and “locates the value of privacy in the development of personal relationships” (Solove 2008: 34–37). Many other possibilities remain, including applying Helen Nissenbaum’s (2004) theory of privacy as “contextual integrity”—a view that focuses on “contextual norms,” especially “norms of appropriateness, and norms of flow or distribution” (119).

Which theory of privacy, if any, is right, goes beyond the scope of this chapter. We also do not address whether it is even necessary to pursue justification for obscurity judgments by selecting only a single theory, and how generalizable obscurity outlooks can be across cultures and regions. The crucial point, here, is that if it is reasonable to maintain that the grandmother should have acquired consent before proceeding further, the grandchildren’s finger-waving must be supported by more ideas than obscurity theory itself can offer.

Obscurity: Probabilistic and Diminishing

A key to understanding obscurity entails appreciating why it is not an absolute safeguard. Obscurity can provide only probabilistic levels of protection by increasing the transactional cost of finding or understanding information. This means securing information through obscurity-promoting activities always involves some level of risk and can never make information wholly inaccessible.

Two causal implications follow from accepting that this is a true empirical state of affairs: Competent and determined data hunters armed with the right tools and motivation—including hackers and government agencies—will be able to find a way to get and decipher obscured disclosures. Less prepared and committed folks, however, will experience the great effort that is required to breach obscurity as a deterrent that influences whether they even try to do so.

While obscurity can be extremely effective in protecting individuals in their day-to-day activities, it is less effective when people have good reasons to dig deep into your data trail. For example, college and job applicants can invite heightened surveillance scrutiny that undermines obscurity protections that in other circumstances work effectively. So can falling in love and out of it—especially if negative motives, like revenge, or adversarial procedures, like divorce, come into play. All it takes is a single trigger event and otherwise strong obscurity protections can come undone. Even if one is lucky enough to avoid unwanted triggers, seven factors are making it increasingly difficult for everyone to obtain and maintain obscurity protections (Hartzog and Selinger 2013c).

First, political attempts to create enhanced accountability through enhanced transparency will lead to more information becoming available that is easy to search. As an illuminating example, consider the controversy that erupted in California when the agency that handles state pensions, CalPERS, announced it was going to make it easy to search for pension information on its database (Hiltzik 2013). While all of the data had long been classified as public information, some retirees felt the change in obscurity would violate their privacy by making it too easy for others to learn what their monthly benefits are.

Second, there is increasing social and professional pressure to share information over digital platforms. For example, if you don’t use social media, you run the risk of being branded antisocial, and, if you don’t use LinkedIn, you’re run of the risk of missing out on good job opportunities. Beyond this convergence of publicity pushing norms and technology, institutions like schools are supporting programs that encourage minimizing obscurity. Consider the Los Angeles school district that hired a firm to monitor students’ public posts on social media

(Martinez 2013). This approach might have the positive effect of reducing bullying, drug use, and truancy. But it might also result in parents and school administrators pressuring kids to keep posts public.

Third, more and more places are using powerful surveillance technologies. Structurally speaking, devices like smartphones and services like YouTube have contributed to the creation of a society where we are always vulnerable to being videotaped.

Fourth, due to advances in big data, it is becoming easier to use computational tools to determine what makes individuals tick by analyzing what people in their social networks say and do. Even if you keep a relatively obscure digital trail, third parties can develop models of your interests, beliefs, and behavior based upon perceived similarities with others who share common demographics. That other people's decisions have such a direct and profound impact on our obscurity means we've entered into an age of "networked privacy" (boyd 2013) and "social surveillance" whereby individual-centered notions of responsibility are becoming increasingly outdated (Marwick 2012).

Fifth, as the social web expands, there is increased pressure for identity verification. Companies like Google are pushing a real-names agenda by linking popular platforms to a single user account. Other companies mandate the use of real names by relying on Facebook and their Connect system to participate with features like comments on stories.

The sixth problem is what scientists who study decision making call "bounded rationality" (Acquisti and Grossklags 2005). Due to cognitive limitations, we humans have a limited ability to predict the full array of potential negative consequences of our actions. But thanks to advances in data aggregation and prediction technology, it is going to be increasingly difficult to predict accurately which disclosures of personal information might be used against you and people that you care about. This problem is compounded by the rise of informational asymmetries: third parties that profit from collecting and analyzing our information—like data brokers—are financially incentivized to create practices that typical consumers and citizens have a hard time understanding.

Finally, data collection, processing, and distribution are going to be increasingly automated. This will dramatically reduce the transactional costs that once served as a protection for finding and understanding hidden and cryptic pieces of personal information.

Obscurity and Contemporary Philosophy of Technology

Within the philosophy of technology, the account that mostly resembles obscurity theory is Luciano Floridi's (2013) ontological interpretation of "informational privacy." For Floridi, informational privacy concerns the freedom to control personal information—to avoid having undesirable parties intrude upon or interfere with it (2013: 230). As he notes, it thus can be conceptually distinguished from other types of privacy: "physical privacy" (bodily intrusions), "mental privacy" (accessing or manipulating minds), and "decisional privacy" (interfering with the process of making important choices) (2013: 229–230).

Floridi argues: "Informational privacy is a function of the ontological friction of the infosphere" (2013: 232). Simply put, this means informational privacy becomes more difficult to protect as factors like infrastructure and technology remove the frictions that make it taxing for unwanted parties to "access" our personal information (2013: 232). To illustrate this idea, Floridi asks us to imagine four students living together in a house: if they have excellent hearing and the walls are thin, little stands in the way of eavesdropping; if the walls are made of thick bricks, however, it becomes harder to overhear personal conversations; and, if the students are deaf, the infrastructure that facilitates or impedes acoustic information will be inconsequential—from an

informational privacy perspective—compared with infrastructure that makes it easier or harder to read lips or fingers communicating with sign language (2013: 231–232).

By framing informational privacy in friction-based terms, Floridi can take issue with those who argue that privacy is radically diminished in the digital age, given how easy it is to record, analyze, store, and share vast amounts of information (2013: 230). Things are more complicated, he argues, because contemporary life: (1) provides a myriad of ways for agents to obfuscate and falsify their information; (2) yields numerous possibilities for agents to erect obscurity barriers that force unwanted parties to overcome friction before they can obtain personal info; and (3) creates new kinds of information eroding agents, like viruses, that introduce new forms of vulnerability into our informational ecosystem.

Crucially, (1) and (2) reveal the untenable abstraction that underlines the “continuist interpretation” of privacy protections diminishing, in an absolute sense, over time (2013: 234). Such abstraction focuses only on the expanding power of selectively chosen technologies that eviscerate friction; it overlooks or ignores technologies and technologically mediated practices that are friction preserving or friction augmenting. With so many variables in play, any bottom-line informational privacy calculus rightfully will be subject to methodological skepticism and accusations of oversimplification. Moreover, with more variables on the horizon that will radically modify “our informational environments, our selves, and our interactions, it would be naive to expect that informational privacy in the future will mean exactly what it meant in the industrial Western world in the middle of the last century” (2013: 237).

Floridi’s account departs from obscurity theory, however, by virtue of two features. First, his discussion of informational privacy is parasitic upon an “informational interpretation of the self” (2013: 2010–2227). His account presents a full-blown metaphysical account of personal identity, and as such goes beyond the more pragmatic and empirical scope of what obscurity theorists address. Second, Floridi argues privacy harms should be construed as an assault on the right to maintain an “inviolable personality,” i.e., to control the information that determines our identity (2013: 244). Since obscurity theory is descriptive and not prescriptive, it can be integrated into Floridi’s normative arguments, but it is also compatible with alternatives that reject Floridi’s core assumptions.

Obscurity and Contemporary Legal Theory

Within contemporary legal theory, the account that most closely resembles obscurity theory is Harry Surden’s (2007) account of “structural rights.”

Surden contends “constraints” that make it hard for others to thwart our interest in limiting access to information safeguard privacy. For example, the law can be understood as a constraint, insofar as punishment can serve as a deterrent. Physical artifacts also can function as constraints, limiting how easy it is for others to obtain our personal information. Fences and locks can enforce property rights and safeguard the contents of our homes. Code can be a constraint, too: if you encrypt a digital diary with state-of-the-art algorithms, only a gifted hacker will be able to gain access.

Surden further maintains that through “latent structural constraints,” technology can make certain behaviors too expensive to engage in on a regular basis, and thus protect our privacy interests by default. For example, when the cost of accessing, aggregating, or analyzing data—say, analyzing a human genome—is extremely high, society can preserve privacy without needing to impose corresponding laws.

While Surden’s account of actual and latent constraints is compatible with obscurity theory, it goes a step further by positing the existence of *structural rights*. Surden argues that if

technological advances reduce transaction costs so much that previously rare, unwanted behavior can become commonplace, the change can deprive citizens of *rights* they once had. He thus urges lawmakers to be vigilant against letting changes in technology change the rights that citizens can rely upon, without adequately reflecting on whether emergent possibilities for violating privacy interests are tantamount to perpetuating legally sanctioned privacy harms.

The key to understanding Surden's thesis is to recognize that in legal theory, discussions occur over different kinds of rights. Legal rights are said to exist when the law positively recognizes entitlements in an authoritative text or interpretation and affords remedy through legal processes when such laws are violated. But these rights—some of which are theorized as 'natural rights'—contrast with normative expectations that people feel very strongly about and rhetorically deem as 'rights,' even though their value-based judgments are not legally proscribed.

Surden thus draws our attention to the possibility that lawmakers can still deem something worthy of legal protection even if they have not assigned it a legal right. In the case of latent privacy interests, for example, we may deal with situations that indicate legal protections were not necessary at a given moment in time because the extant transaction costs reliably inhibited undesirable behavior. For example, the law currently does not prohibit using telepathy to access the thoughts of strangers. But the lack of legal prohibition is not tantamount to an acknowledgement that society embraces mind reading. Rather, there simply is no need to outlaw an action that currently is impossible. Should advances in neuroscience or telepathy ever change the situation, legal enforcement might be necessary.

Obscurity and the Law

For more than thirty years, the concept of obscurity has been tentatively and inconsistently embraced in the law, in spirit if not in name. Legal debates surrounding obscurity can be traced back at least to *U.S. Department of Justice v. Reporters Committee for Freedom of the Press* (1989). There, the U.S. Supreme Court recognized a privacy interest in the "practical obscurity" of information that was technically available to the public, but could only be found by spending a burdensome and unrealistic amount of time and effort in obtaining it.

Unfortunately, since this decision, discussion of obscurity in case law remains sparse. Consequently, the concept remains undertheorized as courts continue their seemingly Sisyphean struggle with finding meaning in the concept of 'privacy.' Still, there is some hope to be gleaned from the fact that notable examples exist where ideas associated with practical obscurity have played an important role in legal considerations.

For example, *Bursac v. Suozzi* (2008), reinforced the value of practical obscurity by placing limits on a DWI 'Wall of Shame.' Although the court recognized that the County Executive's campaign to publicize DWI arrests online served a "legitimate purpose," it cited the *Reporter's Committee* case and expressed concern over giving the public unlimited access to the information in an online format. At issue is the worry that doing so "may affect a legal status and impose specific harm by being available to, inter alia, search engines, credit agencies, landlords and potential employers, for a lifetime, regardless of the underlying outcome of the case" (340).

The fact that too often the U.S. legal system bounds privacy problems to matters where confidentiality and secrecy are violated points to the absence of pervasive obscurity consideration. For example, the limits of Fourth Amendment privacy protection from search and seizure have largely been circumscribed by the "third party doctrine" that essentially states that individuals have no reasonable expectation of privacy in information disclosed to third parties (Solove 2008: 139).

Currently, however, obscurity interests are becoming more pronounced and have found their way into arguments that are chipping away at this doctrine. In her concurring opinion to *United States v. Jones* (2012), Justice Sonia Sotomayor proposed that “it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties” (956–957). She stated that she “would not assume that all information voluntarily disclosed to some member of the public for a limited purpose is, for that reason alone, disintitled to Fourth Amendment protection” (957). Although this not an explicit appeal to obscurity, Justice Sotomayor seems to want to protect information that is disclosed to a small group but remains unknown and, perhaps more importantly, is unlikely to be known to everyone absent some kind of activity directed towards surveillance, aggregation, or publicity.

Obscurity also has come to be reflected in the “mosaic theory” of the Fourth Amendment, a recent interpretation that gets articulated as the approach “by which courts evaluate a collective sequence of government activity as an aggregated whole to consider whether the sequence amounts to a search” (Kerr 2012: 313). Justice Sotomayor, for example, noted that “GPS monitoring generates a precise, comprehensive record of a person’s public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations” (*United States v. Jones* 2012: 955). To be sure, this kind of information is often known to some and disclosed in front of or in a way accessible to others. Yet it also is likely sensitive information, and because the data is not aggregated and accessible in one place, it is obscure.

In the end, it seems that Justice Sotomayor favors obscurity protections, at least in some contexts. She stated:

I would take these attributes of GPS monitoring into account when considering the existence of a reasonable societal expectation of privacy in the sum of one’s public movements. I would ask whether people reasonably expect that their movements will be recorded and aggregated in a manner that enables the Government to ascertain, more or less at will, their political and religious beliefs, sexual habits, and so on.

(957)

Justice Sotomayor is not alone in this opinion; Justice Samuel Alito’s concurring opinion also reflected a tacit embrace of the obscurity theory. He noted that “[i]n the pre-computer age, the greatest protections of privacy were neither Constitutional nor statutory, but practical. Traditional surveillance for any extended period of time was difficult and costly and therefore rarely undertaken” (963). He also stated that “society’s expectation has been that law enforcement agents and others would not—and indeed, in the main, simply could not—secretly monitor and catalogue every single movement of an individual’s car for a very long period” (964). Justice Alito thus emphasizes that obscurity is protected by the inability to continuously monitor individuals, but erodes under long-term surveillance.

Additional obscurity legal challenges loom as courts will increasingly be asked to resolve disputes over automated surveillance technologies like license plate readers and facial recognition technologies as well as disputes over whether privacy can exist in expressed social networks like social media.

Using Code and Infrastructure to Protect Obscurity Online

A particularly important issue for obscurity analysts to consider is the impact of changing how public records are stored and accessed—changes that decrease the frictions involved with searching out personal information. Consider Evgeny Morozov’s (2013) discussion of two paradigmatic cases.

In 2008, a change to the technological landscape made it quite easy to learn that a University of California professor donated \$100 in support of Proposition 8, a ballot initiative to ban same-sex marriage. During this time, Morozov writes:

An enthusiastic computer virtuoso . . . set up Eightmaps.com, a website that takes publicly available donor information . . . and puts it on a Google map, placing a marker—along with the donor’s name and occupation—next to the address.

(2013: 64)

As a result of this disclosure, a chain of events was set in motion that led to the professor becoming “the target of several angry missives,” including one that “was copied to his university colleagues and supervisors” (2013: 64).

Those of us who are political progressives will not be sympathetic to proposals like Proposition 8 that aim to squash civil liberties. However, the main issue here isn’t the rightness or wrongness of a particular political cause: What is at stake is the very process of citizens participating in democratic politics without becoming subject to harassment. That process would be equally jeopardized if a professor contributed the measly sum of \$100 to a left-wing cause and became sharply criticized by right-wing detractors.

As Morozov (2013) observes, this situation is especially fraught because esteem for transparency may be out of sync with how technology has come to be used. When California passed the Political Reform Act of 1974, which stipulated political donors must publicly disclose a host of personal information, different material conditions were operative, and they favored “practical obscurity”:

When the act was passed in the mid-1970s, the lawmakers couldn’t possibly imagine that such information might soon become easily accessible to anyone with a smartphone. The disclosures had to be made ‘public’ in theory but not in practice: the hassle of finding those forms in some dusty, forlorn town hall archives had guaranteed them a life of practical obscurity.

(Morozov 2013: 64)

Now, the example of the politically exposed professor exemplifies the problem of ‘bouncing’: unobscured information migrating to an unanticipated and problematic source. But Morozov’s second case illustrates a different problem: “highlighting and shading” (2013: 73–74). Here, unobscured information distorts impressions and taints reputations by enabling a person’s identity to become synonymous with information whose significance is “blown out of proportion” (2013: 74).

To illustrate, Morozov discusses the plight of SUNY Buffalo law professor James Gardiner, a man whose contributions to different political campaigns were culled from the Federal Election Commission’s (FEC) website and placed on the *Huffington Post*. This act that led to web searches for his name quickly leading to this information (2013: 73–74).

Unlike the California professor in the bouncing example, Gardner was not worried about people criticizing his political choices. Instead, he was concerned that his ability to teach Constitutional Law might be compromised if students learned about his political beliefs and presumed that those ideologies would prejudice his outlook in the classroom.

Drawing from “Campaign Disclosure, Privacy, and Transparency” (2011)—an article co-authored by Deborah Johnson, an ethicist who has made significant contributions to the Philosophy of Technology, Priscilla Regan, a privacy scholar, and Kent Wayland, an anthropologist—Morzov suggests it might be useful to tackle these problems by changing the

format by which certain information is made available. Possibilities include: (1) restricting records of campaign-finance disclosures to a ‘read only’ format, thereby increasing the difficulty of downloading and reproduction; (2) degrading the placement of the FEC website within online searches; (3) employing strategies that “‘tie’ database fields together so that highlighting and shading particular aspects of information by ‘untying’ them becomes problematic”; and (4) binding campaign-finance data to expiration dates that delete the information after a set period of time elapses, possibly having a “self-destruct” code follow the data as it migrates across platforms (2013: 74–75).

Policy makers and companies are increasingly looking to design-based solutions for privacy, in some contexts called *privacy by design*, to ‘bake’ privacy into technologies and services. This often involves an end-to-end embrace of the fair information practice principles (FIPPs) such as notice, choice, data minimization, data security, transparency, use limitations, accountability and access. Design-based solutions could also be used to preserve or create obscurity, including technologies such as privacy settings, smart hyperlinks, and search blockers; search policies in terms-of-use agreements and community guidelines that allow for the use of pseudonyms and prohibit data scraping; and behavioral nudges that make obscurity practices salient and do not attempt to manipulate the user into engaging in obscurity-corrosive practices.

Note

- 1 Some of the material in this chapter appeared previously in contributions we made to *The Atlantic*, *Forbes* and *Slate*. These texts are cited in the References.

References

- Acquisti, A., and Grossklags, J. (2005). Privacy and Rationality. *Security and Privacy, IEEE*, 3(1), 26–33.
- Borgmann, A. (1987). *Technology and the Character of Contemporary Life*. Chicago: University of Chicago Press.
- boyd, d. (2013). Networked Privacy. *Surveillance and Society*, 10(3/4), 348–350.
- Bursac v. Suozzi*, 2008 NY Slip Op 28437 [22 Misc 3d 328] (October 21, 2008).
- Dunbar, R. (1993). Coevolution of Neocortical Size, Group Size and Language. *Behavior and Brain Science*, 16(4), 681–735.
- Dunbar, R. (1998). The Social Brain Hypothesis. *Evolutionary Anthropology*, 6, 178–190.
- Dunbar, R., and Spoor, M. (1995). Social Networks, Support Cliques, and Kinship. *Human Nature*, 6(3), 273–290.
- Floridi, L. (2013). *The Ethics of Information*. Oxford: Oxford University Press.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday Anchor.
- Hartzog, W., and Selinger, E. (2013a). Big Data in Small Hands. *Stanford Law Review Online*, 66, 81 [Online]. Available at: www.stanfordlawreview.org/online/privacy-and-big-data/big-data-small-hands [Accessed April 5, 2014].
- Hartzog, W., and Selinger, E. (2013b). Obscurity: A Better Way to Think About Your Data Than ‘Privacy’. *The Atlantic*, January 17. [Online]. Available at: www.theatlantic.com/technology/archive/2013/01/obscurity-a-better-way-to-think-about-your-data-than-privacy/267283/ [Accessed April 5, 2014].
- Hartzog, W., and Selinger, E. (2013c). The Chilling Implications of Democratizing Big Data: Facebook’s Graph Search Is Only the Beginning. *Forbes* [Online]. Available at: www.forbes.com/sites/privacynotice/2013/10/16/the-chilling-implications-of-democratizing-big-data-facebook-graph-search-is-only-the-beginning/ [Accessed April 5, 2014].
- Hartzog, W., and Selinger, E. (2014). Why Grandman Shouldn’t Have Posted Instagram Pics on Facebook. *Forbes*. [Online]. Available at: www.forbes.com/sites/privacynotice/2014/01/07/why-grandma-shouldnt-have-posted-instagram-pics-on-facebook-2/ [Accessed April 5, 2014].
- Hartzog, W., and Stutzman, F. (2013a). The Case for Online Obscurity. *California Law Review*, 101, 1–45.

- Hartzog, W., and Stutzman, F. (2013b). Obscurity by Design. *Washington Law Review*, 88, 385–418.
- Hiltzik, M. (2013). Should CalPERS Post Pensioners' Financial Data Online?" *Los Angeles Times* [Online]. Available at: <http://articles.latimes.com/2013/jul/19/business/la-fi-hiltzik-20130721> [Accessed April 8, 2014].
- Jarvis, J. (2011). *Public Parts*. New York: Simon and Schuster.
- Kerr, O. (2012). The Mosaic Theory of the Fourth Amendment. *Michigan Law Review*, 111, 311–354.
- Madrigal, A. (2014). Behind the Machine's Back: How Social Media Users Avoid Getting Turned Into Big Data. *The Atlantic* [Online]. Available at: www.theatlantic.com/technology/archive/2014/04/behind-the-machines-back-how-social-media-users-avoid-getting-turned-into-big-data/360416/ [Accessed April 17, 2014].
- Martinez, M. (2013). California School District Hires Firm to Monitor Students' Social Media. *CNN* [Online]. Available at: www.cnn.com/2013/09/14/us/california-schools-monitor-social-media/index.html [Accessed April 8, 2014].
- Marwick, A. (2012). The Public Domain: Social Surveillance in Everyday Life. *Surveillance and Society*, 9(4), 378–393.
- Morozov, E. (2013). *To Save Everything Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.
- Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review*, 79, 119–158.
- Phelps, T. (2014). Anita Hill Reflects on Telling Her Story in the Documentary 'Anita'. *Los Angeles Times*, April 5 [Online]. Available at: www.latimes.com/entertainment/movies/moviesnow/la-et-mn-anita-hill-film-20140405,0,1401467.story#axzz2y3YZrZcG [Accessed April 5, 2014].
- Ricks, V. (2013). Gossip in the Digital Age. In B. Beasley and M. Haney (eds.) *Social Media and the Value of Truth*. Lanham, MD: Lexington Books.
- Solove, D. (2008). *Understanding Privacy*. Cambridge, MA: Harvard University Press.
- Surden, H. (2007). Structural Rights in Privacy. *SMU Law Review*, 60(4), 1605–1629.
- Thaler, T., and Sunstein, R. (2009). *Nudge: Improving Decisions about Wealth, Health, and Happiness*. New York: Penguin Books.
- U.S. Department of Justice v. Reporters Committee for Freedom of the Press* (1989). [Online]. Available at: <http://caselaw.lp.findlaw.com/scripts/getcase.pl?court=us&vol=489&invol=749> [Accessed April 5, 2014].
- United States v. Jones*, 132 S. Ct. 945 (2012).
- Vernon, A. (2014). Dear Amy: Sharing My Grandkids' Instagram Photos [Online]. Available at: www.amyvernon.net/dear-amy/dear-amy-sharing-my-grandkids-instagram-photos-smetiquette/ [Accessed April 8, 2014].
- Westin, A. (1967). *Privacy and Freedom*. London: Bodley Head.

Further Readings

- Anderson, H. R. (2012). The Mythical Right to Obscurity: A Pragmatic Defense of No Privacy in Public. *I/S: The Journal of Law and Policy for the Information Society*, 7, 543–601. (Argues that there is and should not be a right to obscurity in public).
- Green, R. (2013). Petitions, Privacy, and Political Obscurity. *Temple Law Review*, 85, 367–411 (Argues that “political obscurity is an important, unarticulated interest in the current discourse about privacy and petitioning.”).
- Larson, R. G. (2013). Forgetting the First Amendment: How Obscurity-Based Privacy and a Right to Be Forgotten Are Incompatible With Free Speech. *Communication Law & Policy*, 18, 91–119. (Argues that an obscurity-based right to privacy is fundamentally at odds with the First Amendment).
- Strahilevitz, L. (2005). A Social Networks Theory of Privacy. *University of Chicago Law Review*, 72, 919–988. (Argues that social networks theory can help courts determine when obscure information shared outside of social networks can be an actionable violation of privacy).

Copyright Between Economic and Cultural Models of Creativity

Gordon Hull

Is it immoral or illegal to download copyrighted materials off the internet without paying? This question was new when such widespread copying suddenly became available in the late 1990s with a combination of broadband internet in college dorms and facilitating websites like Napster. The music industry made a lot of noise about theft, but the core of their proposition was the entirely plausible economic argument that widespread downloading would put them out of business—particularly since the most downloaded songs were also the ones with the greatest revenue, and that the record labels required the revenue from popular songs to offset their losses on investment in more adventurous and creative, but less remunerative, artists. This hints at something more fundamental: copyright exists to promote the creation and dissemination of creative work. Does the law actually do that? Does it do so as well as it should? The music industry essentially provided one model of how copyright might work, serving as a middleman between artists and their fans. This middleman was necessary because artists lacked the ability to effectively produce (from studio time to equipment to the act of making CDs, production was very expensive) or disseminate (on radio or at music stores) their work. The internet doesn't necessarily discredit this model, but it does force us to notice that this mass-production model of intellectual goods is one of many possible models, and so it forces us to address the fundamental question of how we understand creativity, and how that understanding should be reflected in law. That question is not one that we as a society have been able to answer.

Current U.S. legal theory overwhelmingly treats intellectual property (IP), including copyright, in economic terms, in particular as a legal regime that enables the functioning of a market in intangible goods. Other ways of understanding intellectual property, such as a moral right in the tradition of Lockean or Hegelian theory, have been minority views. The emergence of digital forms of production have made newly apparent a number of the weaknesses in this economic approach, and have led to efforts to understand the production of creative intellectual works differently, often along models derived from critical cultural theory. In what follows, I lay out what I take to be the basic terms of the debate. In the first section, I present the reigning economic model in somewhat reductive terms, and look at its specific failures to account for technologically enabled, nonproprietary forms of creative works, and at what I take to be the cause of this, its inadequate modeling of the creative process. In the second section, I look at the development of an alternative account of authorship and creativity, and indicate why it is at tension with the current economic model. In the final section, I consider the cultural response alongside a more sophisticated economic model, emphasizing their fundamental divergence from one another.

The Blockhead Model of Creativity

James Boswell once memorably said that “no man but a blockhead ever wrote, except for money” (quoted in Tushnet 2009: 517). The standard economic model of copyright in the

United States essentially tries to operationalize this sentiment. More specifically, copyright law is generally understood to attempt to promote the production of literary, musical, and other cultural goods by solving an economic problem in the production of intangible goods like songs or literary texts. The current problems in copyright in turn have a lot to do with the ways that new technologies have changed the nature of that economic problem. Consider first a tangible object like a car. Here, the fixed costs—design, prototyping, building a factory, and so on—are high, but it also costs quite a bit in the form of materials and other resources to build each additional car. Possessing the plans isn't enough, both because building a car requires an entire factory system and because the raw materials that go into a car are themselves expensive. Individuals who want cars thus need to purchase them from car companies, and those producers distribute their fixed costs across an entire production run. Producers have an incentive to invest in producing cars because people who want cars will have to purchase them.

For intangible goods, however, the situation is different. The initial costs of making the good are potentially very high, especially when measured in time. Countless hours can go into writing a novel. Writing a song is only the first step to getting a recording, which may involve paying for studio time and musicians. However, once the initial work is done, the marginal costs of making an additional copy are comparatively very low. As a result, anybody who has access to the work in question has at least the potential incentive to make a copy of it for themselves, if they think that copying it is cheaper than buying it. By banning the making of unauthorized copies, the law adds some incentives against copying, both in the form of most people's desire to obey the law, and in the penalties associated with its violation. Together, these enable "the owner to charge a price for access that exceeds marginal cost" (Posner 2005: 59).

The emergence and widespread dissemination of digital technologies have exposed two kinds of weakness in what one might call, in honor of Boswell, the 'Blockhead' model. One is enforcement: digital media make it very easy for end users to copy for free what they formerly had to pay for. In other words, it turns out that one of the main reasons that people did not reproduce copyrighted work was the difficulty in doing so, not anything about the law (Hull 2003). Indeed, copyright as a legal regime faces a number of problems: most people do not understand it, do not think that much of the activity it forbids either is or should be forbidden, and resent what they take to be the lavish lifestyles of content industry executives. Industry efforts at enforcement are hampered by both the scale of copying and the difficulty in targeting the individuals doing it. Even for those who try to understand copyright, the law is extremely complex and difficult to follow in all but the most obvious cases (Litman 2006). This complexity and its dissonance with cultural norms about sharing erode its moral authority further, which in turn lowers its viability as a property regime (Merrill and Smith 2007). Producers increasingly turn to code-based enforcement strategies like digital rights management (DRM), the equivalent to dealing with speeding by installing speed bumps, rather than increasing police patrols (for a defense, see (Cheng 2006)). Consumers bitterly resent DRM, both because producers use it to do a lot more than stop copying (Elkin-Koren 2002; Hull 2012), and because it appears to take control of devices that they own away from them (Gillespie 2007). In such an environment it is not surprising that, when copying became easy, users stopped paying. This problem has been particularly intractable for the music industry, though the industry's own responses have often been perceived as tone deaf or inappropriate (Knopper 2009).

The second weakness, which has seen much less attention in the popular press, is that there are many forms of creativity that do not seem to fit the Blockhead model. Particularly with the rise of 'Web 2.0' and user-generated content, end users have not just taken the opportunity to copy; they have taken it to create their own content, sometimes by way of newly easy copying. Millions of individually curated blogs and websites form some of this content; this material in particular often relies upon more traditional media to supply inputs, which it then comments

on, reworks, etc. (Balkin 2004). Some sites, such as Wikipedia, are produced by large-scale collaboration of volunteers. The existence of these forms of creativity undermines two of the fundamental assumptions of the Blockhead model: that creativity is work and that people won't work without some sort of market-based incentive structure.

The core insight of a substantial body of recent literature is that the Blockhead model radically underemphasizes both nonpecuniary extrinsic rewards like prestige, and the intrinsic reasons people create. As Yochai Benkler summarizes this research,

for any given culture, there will be some acts that a person would prefer to perform not for money, but for social standing, recognition, and probably, ultimately, instrumental value obtainable only if that person has performed the action through a social, rather than a market, transaction.

(2006: 96)

In other words, the value of some of this material to its creators depends on it *not* being readily explainable in market terms. The internet has facilitated both individual creative works, like blogs and mash-ups, and collective ones, like Wikipedia and the Mars ClickWorkers project (which identifies and maps craters on the planet's surface). For the former, it solves a problem of distribution by making it possible to publish to a potentially indefinitely large audience, nearly for free (whether anyone reads that material is a separate consideration). For the latter, the internet solves a problem of scalability. Historically, non-market-based forms of production have been limited to small, tightly knit social communities because it was very difficult to coordinate larger groups of people. For projects that are sufficiently modular (divisible into discrete tasks) and granular (those tasks can be performed by individuals, often in their spare time), the internet not only allows collaborators to self-select from an enormous number of individuals, who contribute to the extent they want or are able, but also to efficiently distribute those contributions. The result is "the emergence of information production that is not based on exclusive proprietary claims, not aimed toward sales in a market for either motivation or information, and not organized around property and contract claims to form firms or market exchanges" (Benkler 2006: 105).

Benkler's formulation is not without difficulties. First, as he notes, not all kinds of work are easily modular or granular. Encyclopedias make sense on this model; novels do not. Second, as Lior Strahilevitz notes, one of the main advantages of social production is the ability to efficiently recruit individuals to work on projects. This also poses a problem of resilience to malicious users. Wikipedia's difficulties in dealing with malicious edits and sock puppetry are examples. Another is the violent harassment faced by women who attempt to participate online (Strahilevitz 2007: 1485, 1493; for harassment, see Citron 2014). Third, large corporations like Facebook and Google emerge to capture market value from users' creative outputs (Benkler 2006). This raises Marxist concerns about the exploitation of labor and the worry that the internet can easily function as a giant vehicle for surplus-value extraction (Terranova 2000). However these larger issues play out, however, it seems to be the case that social production provides a non-IP-based form of innovation.

Benkler's work intersects with a complementary body of research, most prominently by Kal Raustiala and Christopher Sprigman, on highly innovative sectors of the economy that do not rely on IP protection. Raustiala and Sprigman's initial case study (2006) is of the fashion industry. Clothing falls through the cracks in copyright law: because garments are useful, copyright applies only when the 'expressive' components can be separated from the useful ones. Since things like sleeve length, hemline, and fabric stylization cannot generally be removed from the garment, they are not protected by copyright (similar obstacles are present for trademark and

patent protection). In this space of low-IP protection, however, flourishes a large and extremely creative industry. Raustiala and Sprigman argue that this industry is innovative not because of the presence of IP, but because of the lack of it: because fashion is a positional good (I want to be more fashionable than those around me), the arrival of knock-off and imitation designs motivate leading designers. The fashion cycle, in other words, is driven by copying: when a given look becomes mainstream, due to copying, the fashionable demand something new. Without that copying, creativity in fashion would grind to a halt. Subsequent research (see the summary in Rosenblatt 2011) has discovered other such ‘low-IP’ parts of the economy. Some, like fashion, are enabled by doctrinal quirks. Others, like academic writing, tattoos, and sports moves, exist in cultures that manage creativity with non-IP-based norms. Still others, like parody, exist in the space of specific legal carve-outs. But they all challenge the necessity of making the assumptions undergirding the Blockhead model’s incentives narrative.

In a more theoretical vein, scholars such as Roberta Kwall and Rebecca Tushnet have pointed to fundamental difficulties caused by the incentives argument’s reliance on a reductive, behaviorist account of the reasons people do things: as Tushnet puts it, it is as if “the scientist might have been a journalist, if only her internal utility calculus and/or the relative rewards from the two fields differed enough.” The problem is that “the actors in this story are unrecognizable as people. Creativity, as lived, is more than a response to incentives, working from fixed and random preferences” (Tushnet 2009: 521). Tushnet cites several published accounts by authors about their own motivations, which often do not seem susceptible to economic explanation. Kwall (2006) highlights the importance that authors assign to “inspiration” and the intrinsic value they find in expressing themselves in their works; these values are also deeply embedded in religious accounts of divine and human creation. Feminist scholarship has looked particularly at fan fiction: produced largely by women, and often exploring issues of sexuality that are difficult to explore in more commercial contexts, fan fiction is a form of expression that seems largely indifferent to the legal modeling of creativity in copyright (Katyal 2006a; Tushnet 2009).

This research speaks primarily to the motivations of individual creators. In the case of corporate creators or media companies, the Blockhead model works much better, as companies have to pay whomever they hire to do creative work, and they expect that the work will contribute in some way to corporate profitability. However, recent work by Jessica Silbey (2015) complicates even this story considerably. On the one hand, she adds further evidence that individual creators are not monolithically motivated by the incentive structures of copyright. On the other hand, pointing out that it is often a mistake to treat large corporate entities as univocal actors, she reports on a series of interviews with corporate agents, whose “descriptions of corporate culture and management choices reject the monolithic identity of a corporation whose only goal is to make money,” and “even for the companies that may exist to make money . . . the progress of science and art requires passion for the work, which does not appear to be incentivized by IP’s investment function.” When IP law does intervene, it often “serves goals relating . . . to relationship building and business flexibility” (2015: 13–14). In sum, the economic incentives model is important. However, such incentives seem to be neither necessary nor sufficient to explain why people create more broadly.

Incentives, Authorship, and Property

The emergence and widespread adoption of digital technologies, then, has served to substantially disrupt the legal account of creativity underpinning the copyright regime. By making it easier to create (and easier to see the creations of those who create) in ways not well-modeled by the Blockhead theory, digital technologies have thrown into question the relation between

creation and markets. Cultural phenomena like peer-to-peer or other distributed forms of ‘commons-based’ production in particular raise the challenge of understanding the emergence of order and innovation without the intervention of markets. At the same time, an emergent body of research raises the question, not of order without markets, but of functioning, innovative markets in immaterial goods that do not rely on intellectual property’s understanding of incentives. Finally, a growing body of empirical research indicates that many individual actors create for reasons not captured by the Blockhead model.

In this section, I suggest that the disruption to the economic account of creativity is in part evidence of a disruption to the basic social roles the copyright regime envisions. One very important role for legal and other rule-based systems for the function of markets is their ability to define appropriate social roles for agents in market interactions and to enforce appropriate behaviors (such as contract performance) for those in such social roles. Indeed, without these sets of rules defining social roles, the predictability and stability that enables markets to function would not exist (see Lewis 2014: 9 of 24, citing Hayek). The basic social roles contemplated by copyright are those of author and consumer. It is the disruption of these roles that underlies the disruption in the regime’s model of incentivized creation.

Many encounters with cultural goods have always been outside of the market-defined role of consumers. Even in the age of mass media, a lot of cultural activity—from singing copyrighted songs together to passing along copies of used books—happened outside the realm of markets. Much of it was either tolerated by copyright owners or protected as fair use, in part because of the difficulties in monitoring, measuring and monetizing these social exchanges (Lessig 2006: 185–191). Digital technology changes this pattern in several ways. First, since nearly all digital use involves making a copy, engaging culture increasingly requires engaging the legal system that defines use as consumption. Second, the availability of inexpensive editing technologies has encouraged people to experience cultural goods interactively, and to create by using existing works as raw materials for mash-up, parody, and other forms of reworking. Third, digital methods of distribution make such nonmarket uses much more potentially disruptive to market uses: making a mix tape for a friend is fundamentally less market-disruptive than putting the same material in a shared folder on LimeWire. Finally, in response to these developments, the legacy-creative industries have tried very hard to push cultural interaction into market-given norms of paid, passive consumption.

In response, many scholars point to reasons why the market model of consumption is a poor fit for cultural interactions, even if those interactions could, in theory, be modeled economically. One strand of arguments highlights the extent to which our encounters with cultural goods are vital components of our individual and collective identities. For example, situating his analysis in an account of Hegel, John Tehranian notes that we develop and express our identities through the commodified properties that fill our environment. As a result, a move from tangible goods to intangible ones covered by intellectual property laws puts sudden constraints on identity formation: “one can contextualize and communicate one’s relationship with one’s jeans by wearing them in public, but the equivalent act of publicly utilizing a copyrighted work would impinge on an author’s exclusive right to control public displays and performances” (Tehranian 2011: 27). Thus, copying an advertising slogan onto one’s webpage violates IP law, but wearing the slogan on a licensed shirt does not. Further, as Sonia Katyal notes, cultural phenomena like fan fiction underscore that the meaning of works is not something passively absorbed by consumers; rather, it emerges in a space between authors, performers, and audience, in a way that none fully control (Katyal 2006a). More generally, as Madhavi Sunder proposed, “rather than narrowly viewing intellectual property as *incentives-for-creation*, we must understand intellectual property as *social and cultural relations*. Intellectual property rights structure social relationships and enable, or disable, human flourishing” (2006: 285). Although one can argue

that human flourishing can be completely captured in market terms, a wide range of theorists (such as Martha Nussbaum and Amartya Sen, from whom Sunder is drawing here) provide other, broader models of what makes a human life a good one.

Somewhat less frequently discussed, but equally important, is the law's understanding of the 'author' as the originating locus of intellectual property. It is the author who is supposed to provide the consumer with the goods they seek, and many of the difficulties in the Blockhead account can be traced to individuals who imperfectly fit its authorial role. In his landmark *Shamans, Software and Spleens*, James Boyle (1997) suggested that this dominant economic discourse was problematically suffused with a "romantic vision of authorship" that viewed creation as the product of an isolated creator whose work is both original and transformative. Boyle refers to Foucault (1984), who had proposed that authorship, whatever its metaphysical status, serves an important political function, one that our cultural fixation on authorship needs to address clearly. We use the author-function, he proposes, to organize our texts and writing, associating them with particular individuals and not with others (Foucault 1984: 107). Or, as Laura Heymann (2005) puts it (with reference to Foucault), designations of authorship have a trademarking function as they give the work a brand identity. Legally, the author-function enables the designation of a single creator, whose works (or at least contributions to works) can be individuated from the efforts of others; they can then be remunerated, thereby incentivizing the author's creative activities. In other words, if the Blockhead model flattens our interactions with cultural goods into passive consumption, it similarly misrepresents how those goods are produced.

As Boyle (1997) noted, basing our legal regime on this understanding of authorship has significant consequences for non-Western cultural forms, as 'traditional' cultural products often do not have identifiable individual authors, and often do not exist in discrete works. This raises the immediate possibility of biopiracy and other forms of exploitation of these cultural forms that appear to be 'unowned.' For example, the rock group Enigma's "Return to Innocence," which made millions of dollars for the band and which was used in television ads for the 1996 Summer Olympics, turns out to have been lifted from the indigenous Ami people of Taiwan, who saw their sacred music exploited and who received nothing in return (Riley 2000).

Even within Western societies, however, the model functions awkwardly because authorship can be—often deliberately—difficult to measure in discrete quanta. Different recording tracks can be separated and combined in mash-ups like "The Grey Album" (Sunder 2006: 303). Music can be combined with video to create social commentary, as in Sloane's "Star Trek Dance Floor," which combined footage from the *Star Trek: Reboot* movie with the song "Too Many Dicks on the Dance Floor" by Flight of the Conchords to critique the machismo (and lack of women) in the movie. Artworks are often in communication—implicit or explicit—with other art forms, and almost any work will contain a lot of borrowing, conscious or otherwise. For example, when Kanye West stood on national TV and, gazing at the ruins of immediate post-Katrina New Orleans (and the lack of an adequate federal response to the situation), proclaimed that "George Bush doesn't care about Black people," New Orleans-based The Legendary K.O. produced a song of the same name, borrowing liberally from West's "Gold Digger." West's song, in turn, refers to and borrows from Ray Charles's "I Got a Woman." Charles himself took liberally from the gospel standard "I've Got a Savior." The legal status of "I've Got a Savior" was unclear, and nobody asked. Under the different and more expansive copyright laws of today, however, Boyle notes, "one thing is clear. Much of what Charles and [Renald] Richard did in creating their song would be illegal" (2008: 137). So too, rap sampling has been essentially shut down as a matter of law following a court decision that ruled that NWA's sampling and looping (into unrecognizability) of a short guitar riff constituted infringement (*Bridgeport Music v. Dimension Films* 410 F.3d 792 [6th Cir. 2005]).

In some cases, creative expression arguably consists in repeating verbatim someone else's expression (Tushnet 2004). The expression thus is the copying, but does not reduce to the copied work. For example, people often feel that a particular song expresses perfectly an aspect of their own lives. Individual performances of works have different effects, depending on who performs them to what audience. Editing and selecting what to include in compilations is an expressive act. Political campaigns have theme songs. Sometimes, the meaning of the expression would be lost entirely if it weren't copied directly: when Martin Luther King, Jr., delivered the "I Have a Dream" speech, the phrase was unique neither to him nor that speech. However,

What was historically significant was that huge numbers of people in the audience had never heard him preach. For them, that he was copying was irrelevant to the power of his words. It didn't matter that the expression wasn't original; it did matter that it was King speaking, at that time, at that place, to those people.

(Tushnet 2004: 576)

Those who repeat King's words want to associate themselves with the Civil Rights movement, just like those who repeat the Pledge of Allegiance want to associate themselves with a particular understanding of patriotism. Repeating only part of the Pledge wouldn't convey the same thing, and rewriting it would convey even less. Religion often requires the exact copying of material that is said to be sacred in origin. In one notable case, a court enjoined a splinter sect from copying religious texts owned by the parent organization (*Worldwide Church of God v. Philadelphia Church of God*, 227 F.3d 1110 [9th Cir. 2000]).

Cases of expressive copying are extreme examples, but they underscore that the irreducibly social nature of authorship and expression, and the demands of the author-function exist in uneasy tension. One way to put the point is that property regimes require boundaries and modularity, and one of their functions is to conceptualize and designate identities that can then be treated as commodities on the market (Fennell 2012). For intangible goods, neither production nor consumption is necessarily 'lumpy.' Borrowing, homage, and so forth are very hard to quantify, and efforts to do so often generate troubling results. On the consumption side, there are no necessary threshold moments that have to be reached for material to be enjoyable (you don't have to see the whole movie, or listen to a song five times to enjoy it). Copyright has traditionally allowed a single purchase to cover largely unrestricted and unregulated subsequent use, including selling the good to a third party. The mediation of creative activities by digital technologies introduces the capacity to attempt to treat consumption as lumpy by metering use, requiring payment for each access to a work, and using licensing terms and DRM to prevent sale to third parties. At the same time, it encourages both works and kinds of work that not only emphasize the presence of the many different contributions to a given work, but also the difficulty in neatly distinguishing them. In other words, they tend to make all work look less like the original creation of a romantic authorial genius and more like the Taiwanese Ami's sacred music.

Economy or Culture?

Although the sort of work outlined earlier offers an alternative account to a simple incentives-based model of IP, it should be stressed that economic models can also respond to some of the problems in that simple model. Boyle's work presented an early example of one response to the difficulties in standard economic accounts of copyright law, one that attempted to think about culture more broadly than incentive-based economic models, and that was particularly attuned to social justice issues that those models can occlude. A review of Boyle's book by

Mark Lemley (1997) presented another kind of response. For Lemley, the problem was not so much a legal misunderstanding of authorship and creativity as an excessive focus on proprietization. In other words, the problem with the economic model is that it is bad economics; the solution is better economics.

Neoclassical economic theory generally supposes that the reason to institute property regimes (as opposed to commons-based or other forms of cultural regulation) is to resolve resource allocation problems. Consider a field used for grazing. One way of allocating access to the field would be as a ‘commons’: everyone in the village is allowed to use it freely. The problem is that no one has an incentive to use the resource sustainably, since no one bears the cost of their own overuse. As a result, it is inevitably overgrazed in a ‘tragedy of the commons.’ The solution is to allocate parcels to individual shepherds. Each shepherd then has the incentive to manage their own parcel sustainably, since they suffer the full cost of pollution, depletion, and such. The harms—‘negative externalities’—are avoided by making them directly harmful to the individuals who cause them, in a process of ‘internalizing’ them.

As noted in the first section, the case for intangible goods is related, and is based on the supposition that inventors need to be able to realize the social benefits associated with their creations, since capturing this value is what provides them adequate incentives to create. For example, if a song brings joy to millions of people, the writer needs to be able to capture some of that value. As Lemley and Brett Frischmann describe the theory:

The basic idea behind ‘internalizing externalities’ is that if property owners are both fully encumbered with potential third-party costs and entitled to completely appropriate potential third-party benefits, their interests will align with the interests of society, and they will make efficient (social welfare-maximizing) decisions.

(Frischmann and Lemley 2006: 109)

Lemley (2005) pointed out that the focus on proprietization in IP seemed to follow this logic, but that it shouldn’t do so without serious revision: for intangible goods, the externalities are generally positive, and creators only need to capture *enough* value to have an incentive to create. Giving them too many property rights becomes socially inefficient because there are harms to society caused by overproprietization. In other words, there is no reason to think that adding more reward always causes more innovation, as the Blockhead model seems to imply.

Current copyright law incorporates this point to some extent. For example, ‘ideas’ cannot be copyrighted—only their ‘expressions.’ *Romeo and Juliet* could be copyrighted, but not “play about star-crossed lovers.” As R. Polk Wagner (2003) notes, this means that copyrighted shows like *Survivor* simultaneously create noncopyrightable ideas like ‘reality TV,’ ideas that then potentially generate a lot of follow-on, copyrightable content. So too, ‘fair use’ can be used to allow socially beneficial uses of material, like critique, that creators might not like. Copyright terms are also limited: at some point, copyrighted material inevitably passes into the ‘public domain,’ free for anyone to use. Lemley and Frischmann and other economically oriented scholars share with more culturally oriented theorists the concern that these protections are no longer robust enough.

They differ sharply, however, in what to do about the situation. For Lemley and Frischmann, copyright protections should be better aligned to the promotion of positive externalities (‘spillovers’), because of the great social value those generate. For example, the 1998 retroactive extension of copyright by Congress (which had the effect of keeping Mickey Mouse protected) made no economic sense at all, and most of the leading neoclassical economists signed onto an amicus brief opposing it when it was challenged (in *Eldred v. Reno*, the Supreme Court upheld the statute on noneconomic grounds). So too, Robert Merges (1993) has argued that current

protections for parody are too narrow, because they require that the work being used is also the one being parodied—so, for example, the use of Mickey Mouse to parody 1950s cultural norms is unprotected, but the retelling of *Gone With the Wind* from a slave’s point of view (Alice Randall’s *The Wind Done Gone*) is protected. Given that certain characters and works (like Disney movies) are emblematic of cultural norms, surely those should be protected too? Economic reasons can also mitigate fears of overenforcement: Steven Hetcher (2009) argues that fan fiction does not need further protection now, because social norms favoring remix culture also create significant costs (e.g., bad PR) for copyright owners who would try to enforce their rights.

As suggested earlier, for cultural theorists, the reasons to protect cultural expression are normative for reasons that do not directly reduce to welfare economics, but instead are embedded in broader concepts of human flourishing. Katyal (2006b) argues that parody and critique, for example, are good for democracy, and such expression deserves more legal protection than it gets. More generally, a broader notion of innovation and creativity allow the law more fully to protect the many kinds of cultural expression that are not captured by the current legal regime and that are not reducible to current, economistic models of social welfare. As Sunder notes, “we must understand intellectual property as social and cultural policy” (2012: 32). It is not that economics should be eliminated—far from it—but that such analysis is incomplete as cultural policy, since it neglects “information’s role in cultural life and human flourishing” (2012: 31). A theory that reduces culture to economic incentives for that reason needs to be revised: it tends to reduce the totality of human culture to what can be measured in markets and market transactions. Following Amartya Sen and Martha Nussbaum, Sunder (2012) underscores that there are many ways of measuring human flourishing, and that welfare economics can miss many of them. As she put it in an earlier paper, one particularly needs to attend to “the crucial question . . . who will have the power to make our cultural world” (Sunder 2006: 320), and whether an incentives-based, market-oriented property regime is the only or even the best way to answer the question.

One area where the theories diverge is over what economic theory calls ‘deadweight loss.’ Because copyright assigns a monopoly on a given good, it allows the good’s owner to sell it at a monopoly price. This is necessary: since the marginal cost of reproducing the good is near zero, almost any price would be above a natural market price, which would approach zero. Deadweight loss refers to the inevitable existence of users who value the good, but either cannot or will not pay the monopoly price. Allocating social resources on the basis of market price demonstrably favors those with greater ability to pay (Baker 1975); Benkler notes that, in the case of IP, those who cannot pay are often those whose uses are highly socially valuable, such as those pursuing basic science (Benkler 2000). The divergence is over what to do about this. As Boyle (2000) notes, the economic solution to deadweight loss is price discrimination: the accumulation of data about individual consumers so that they can be charged the price that they are willing/able to pay, combined with a regime that makes it difficult for those willing to pay more to obtain the product for less. Digital technologies—and certainly the capabilities of big data are currently already far beyond what Boyle imagined in 2000—makes very good price discrimination possible. The question is whether we, as a society, are comfortable with that level of market saturation in our lives, especially given that big data introduces significant worries about privacy, discrimination and opacity in governance (see, e.g., boyd and Crawford 2012; Pasquale 2015; Tufekci 2014). Cultural theorists see price discrimination as misguided because it solves the wrong problem: much of culture should exist outside market relations, even if we have the technical capacity to subsume it into markets. An excessive focus on markets makes us forget why the spaces outside of markets matter; or, as Michael Hardt and Antonio Negri memorably put it, “private property makes us stupid” (Hardt and Negri 2004: 188).

Conclusion

Copyright, by design, exists at the intersection of market economics and cultural expression. There are relatively few who would argue that economic analysis has no place in copyright law. However, the emergence of digital production and the increasing importance of intangible goods have made salient both the difficulties in the Blockhead model's understanding of creativity and the extension of intellectual property into forms of cultural expression formerly either outside of it, either by default or by design. In other words, digitality is increasing the importance of IP to culture, and so both the question of how IP understands culture and creativity, and the underlying question of where one should properly draw the boundaries of economic analysis, emerge as vital philosophical questions underlying a great deal of the legal debates around copyright and its proper scope.

References

- Baker, C. E. (1975). The Ideology of the Economic Analysis of Law. *Philosophy and Public Affairs*, 5, 3–48.
- Balkin, J. M. (2004). Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society. *New York University Law Review*, 79(1), 1–55.
- Benkler, Y. (2000). An Unhurried View of Private Ordering in Information Transactions. *Vanderbilt Law Review*, 53, 2063–2080. doi:10.1162/DAED_a_00121
- Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.
- boyd, d., and Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. doi:10.1080/1369118x.2012.678878
- Boyle, J. (1997). *Shamans, Software and Spleens: Law and the Construction of the Information Society*. Cambridge, MA: Harvard University Press.
- Boyle, J. (2000). Cruel, Mean, or Lavish? Economic Analysis, Price Discrimination, and Digital Intellectual Property. *Vanderbilt Law Review*, 53, 2007–2039.
- Boyle, J. (2008). *The Public Domain: Enclosing the Commons of the Mind*. New Haven, CT: Yale University Press.
- Cheng, E. K. (2006). Structural Laws and the Puzzle of Regulating Behavior. *Northwestern University Law Review*, 100, 655–717.
- Citron, D. K. (2014). *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.
- Elkin-Koren, N. (2002). It's All About Control: Rethinking Copyright in the New Information Landscape. In N. Elkin-Koren and N. W. Netanel (eds.) *The Commodification of Information*. The Hague: Kluwer, 79–106.
- Fennell, L. A. (2012). Lumpy Property. *University of Pennsylvania Law Review*, 160, 1955–1993.
- Foucault, M. (1984). What Is an Author? In P. Rabinow (ed.) *The Foucault Reader*. New York: Pantheon Books, 101–120.
- Frischmann, B. M., & Lemley, M. A. (2006). Spillovers. *Columbia Law Review*, 100, 101–143.
- Gillespie, T. (2007). *Wired Shut: Copyright and the Shape of Digital Culture*. Cambridge, MA: MIT Press.
- Hardt, M., and Negri, A. (2004). *Multitude: War and Democracy in the Age of Empire*. New York: Penguin.
- Hetcher, S. A. (2009). Using Social Norms to Regulate Fan Fiction and Remix Culture. *University of Pennsylvania Law Review*, 157, 1869–1935.
- Heymann, L. A. (2005). The Birth of the Authornym: Authorship, Pseudonymity, and Trademark Law. *Notre Dame Law Review*, 80, 1377–1449.
- Hull, G. (2003). Digital Copyright and the Possibility of Pure Law. *qui parle*, 14, 21–47.
- Hull, G. (2012). Coding the Dictatorship of 'The They': A Phenomenological Critique of Digital Rights Management. In M. Sanders and J. J. Wisniewski (eds.) *Ethics and Phenomenology*. Lanham, MD: Lexington Books.

- Katyal, S. K. (2006a). Performance, Property, and the Slashing of Gender in Fan Fiction. *Journal of Gender, Social Policy and the Law*, 14, 461–518.
- Katyal, S. K. (2006b). Semiotic Disobedience. *Washington University Law Review*, 84, 489–571.
- Knopper, S. (2009). *Appetite for Self-Destruction: The Spectacular Crash of the Record Industry in the Digital Age*. New York: Free Press.
- Kwall, R. R. (2006). Inspiration and Innovation: The Intrinsic Dimension of the Artistic Soul. *Notre Dame Law Review*, 81, 1945–2012.
- Lemley, M. A. (1997). Romantic Authorship and the Rhetoric of Property. *Texas Law Review*, 75, 873–906.
- Lemley, M. A. (2005). Property, Intellectual Property, and Free Riding. *Texas Law Review*, 83, 1031–1104.
- Lessig, L. (2006). *Code and other Laws of Cyberspace, Version 2.0*. New York: Basic Books.
- Lewis, P. (2014). Notions of Order and Process in Hayek: The Significance of Emergence. *Cambridge Journal of Economics* 39(4), 1167–1190. doi:10.1093/cje/beu043
- Litman, J. (2006). *Digital Copyright*. Amherst, NY: Prometheus Books.
- Merges, R. P. (1993). Are You Making Fun of Me? Notes on Market Failure and the Parody Defense in Copyright. *American Intellectual Property Law Association Quarterly Journal*, 21, 305–312.
- Merrill, T. W., and Smith, H. E. (2007). The Morality of Property. *William and Mary Law Review*, 48, 1849–1895.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- Posner, R. A. (2005). Intellectual Property: The Law and Economics Approach. *Journal of Economic Perspectives*, 19(2), 57–73.
- Raustiala, K., and Sprigman, C. (2006). The Piracy Paradox: Innovation and Intellectual Property in Fashion Design. *Virginia Law Review*, 92, 1688–1777.
- Riley, A. R. (2000). Recovering Collectivity: Group Rights to Intellectual Property in Indigenous Communities. *Cardozo Arts and Entertainment Law Journal*, 18, 175–225.
- Rosenblatt, E. L. (2011). A Theory of IP's Negative Space. *Columbia Journal of Law and the Arts*, 34, 317–365.
- Silbey, J. (2015). *The Eureka Myth: Creators, Innovators, and Everyday Intellectual Property*. Stanford, CA: Stanford Law Books, an imprint of Stanford University Press.
- Strahilevitz, L. J. (2007). Wealth Without Markets. *Yale Law Journal*, 116, 1472–1516.
- Sunder, M. (2006). IP3. *Stanford Law Review*, 59, 257–332.
- Sunder, M. (2012). *From Goods to a Good Life: Intellectual Property and Global Justice*. New Haven, CT: Yale University Press.
- Tehrani, J. (2011). Parchment, Pixels and Personhood: User Rights and the IP (Identity Politics) of IP (Intellectual Property). *Colorado Law Review*, 82, 1–84.
- Terranova, T. (2000). Free Labor: Producing Culture for the Digital Economy. *Social Text*, 18(2), 33–58.
- Tufekci, Z. (2014). Engineering the Public: Big Data, Surveillance and Computational Politics. *First Monday*, 19(7). doi:10.5210/fm.v19i7.4901.
- Tushnet, R. (2004). Copy This Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It. *Yale Law Journal*, 114, 535–590.
- Tushnet, R. (2009). Economies of Desire: Fair Use and Marketplace Assumptions. *William and Mary Law Review*, 51, 513–546.
- Wagner, R. P. (2003). Information Wants to Be Free: Intellectual Property and the Mythologies of Control. *Columbia Law Review*, 103, 995–1034.

Cyberwarfare

Benjamin C. Jantzen

Introduction

Imagine a team of commandos preparing to strike at an unsuspecting nation-state. They carry no obvious weapons. They are not dressed in fatigues. In fact, they wear civilian clothing and occupy a nondescript office building on a suburban street. In the operation about to take place, they will not breach any national borders, at least not in person. They will not infiltrate enemy installations or embassies. Instead they attack by sitting at their desks with a radio transmitter connected to a laptop. The transmitter talks the language of the ‘smart meter’ that the electric company has installed outside their building. The simple computer called a *microcontroller* in each meter allows it to communicate through a hierarchy of progressively more complex computers to a central control facility of the company that owns the equipment. When the commandos’ laptop engages the meter, it exploits a vulnerability in its modest operating system to implant a block of malicious code. This code subtly alters the behavior of the meter, causing it to broadcast to other meters within range and infect them with the same block of code. The malicious code rapidly spreads from meter to meter across a large geographic region. At the same time, each infected meter passes tainted data to the control node above it, until finally the upward cascade ends with the implantation of a payload on the relatively powerful computers of whatever control center that particular meter reports to. At first nothing happens; the weapon has yet to find its target. But eventually, the implanted code leaps the right number of national borders as the number of infected meters grows exponentially. At some point thereafter, the infection spreads upward to infect the control computers of the targeted company. There, the foreign code surreptitiously opens a communication channel to the commandos, and signals them that it’s open for business. A much larger piece of code is then uploaded through the hijacked computers of some innocent bystanders. This much larger digital weapon sets to work coordinating the final stage of the attack. The corrupted control software incites the targeted company’s control system to direct a surge of power to a carefully selected set of substations, all at precisely the same moment. At each substation, the surge induces the large transformers at the heart of the station to explode in spectacular fashion, leaving the industrial operation serviced by that particular substation without power from the grid. At each point of attack, diesel generators then roar to life so that critical (and dangerous) manufacturing processes can be safely halted until main power is restored. But the control systems of these generators—also equipped with microcontrollers running tiny programs—have already been corrupted by the commandos. Rather than maintaining a steady flow of electricity, each diesel generator wildly oscillates the throttle, and is quickly reduced to a smoking hulk as the combustion engines tear themselves apart. There is no way to control the manufacturing process that is now running amok. At that point, three munitions plants spread across a country that does not believe itself

to be at war simultaneously and—to anyone but the digital commandos—mysteriously explode, resulting in significant loss of life, matériel, and war-fighting capability.

This sort of scenario—or worse—is what those most exercised by the prospect of ‘cyberwarfare’ fear. While nothing quite like this has ever taken place, it is within the realm of practical possibility as I write. The possibility of infecting wireless electrical meters with self-propagating, malicious code has already been demonstrated (Giannetsos et al. 2010; Naone 2009; Zetter 2015a: 155–157). The infamous Stuxnet—a digital weapon designed and deployed in a collaborative effort by the United States and Israel to slow the advance of Iran’s nuclear program—demonstrated the very real plausibility of infecting multiple layers (albeit from the top down) of a remote hierarchical control system like that found in industrial applications throughout the developed world, including systems that run American and European power grids (Zetter 2015a). Malicious code has already opened a back door into regional electrical supplier control systems that was recently used by hackers to cause blackouts in Ukraine (Goodin 2016). The destruction of diesel backup generators by remotely injected computer code was demonstrated for reporters as part of Project Aurora (Meserve 2007a, 2007b). And the fact that many chemical manufacturing processes cannot be halted abruptly without dire consequences is hardly a secret (U.S. Environmental Protection Agency 2001).

Such a malicious manipulation of computing technology is possible. But would it be an act of war? If so, is it more of the same as far as innovation in warfighting goes, or do such attacks reflect a use of force that is of a kind genuinely new under the sun? How much risk does cyberwarfare really pose? Does it raise new ethical questions? Despite the hype surrounding the term *cyberwarfare* there exists no clear explication of the concept, and consequently, no consensus on these questions. In fact, there isn’t even a consensus on how to spell the term—*cyber war*, *cyber-war*, and *cyberwar* are all contenders.

It is the aim of this chapter to explicate the phenomenon of cyberwarfare itself. In other words, it’s my goal to produce a clear definition of cyberwarfare that captures all or most of the cases on which there is agreement while simultaneously exposing the characteristic features shared by all these cases. Such an explication makes it plain what, if anything, sets cyberwarfare apart from other modes of conflict, and lays the groundwork essential for tackling the harder questions of risk and morality. In order to accomplish this aim, I’m going to follow an indirect strategy. Though an appropriate definition of cyberwar is controversial, many activities are more or less uncontroversially acknowledged to be ‘cyberattacks.’ So I will proceed by explicating the notion of a cyberattack. Then, assuming that warfare essentially involves attacks of one sort or another by one nation-state upon another, we can understand *cyberwarfare* as warfare involving cyberattacks.

It’s important to note at the outset that there is a tendency to use terms like “attack” and “war” rather promiscuously to label an enormous and heterogeneous collection of events. But, as James Lewis points out, “it’s unhelpful and incorrect to call every bad thing that happens on the internet a ‘war’ or ‘attack’” (2011: 23). So I’m going to ignore activities that clearly fall under the rubric of espionage, crime, or activism. Specifically, I’ll avoid discussion of such issues as cyberespionage, ‘hacktivism,’ and cyberterrorism.

What Is a Cyberattack?

So what then is a cyberattack? In this section, I present some examples described along three dimensions: by their mode of influence (i.e., how they affect their computational targets), by the means of influence (i.e., how the attack reaches its target), and by the aims of influence (i.e., what the attack targets or is intended to accomplish). This scheme allows for easy generalization beyond the contingent collection of modern technologies and methods.

Mode of Influence

Resource Attacks

In a resource attack, the attacker aims to adversely influence a target system by depleting its available communication resources. An old (in internet years) and common kind of resource attack is the Denial of Service (DoS). All DoS attacks work by flooding the communication ports of a target machine with packets of data. There are many variations on this theme, but one prominent method is the Distributed DoS (DDoS) attack. In a DDoS attack, many computers from different locations send connection requests in rapid succession and swamp a web server so that no traffic can get through, rather like the Three Stooges trying to pile through a door together. It is a way of ‘denying service’ to the normal users of the website or service hosted by the server. Because the computers engaged in a DDoS are often infected by malicious software that unites them, without the knowledge of their owners, into a ‘botnet’ under the control of a distant ‘bot-herder,’ the number of attacking machines can be truly staggering. The largest known botnet, Conficker, encompassed some nine *million* compromised machines worldwide (“Clock ticking on worm attack code,” 2009).

Configuration Exploits

Other attacks exploit properly functioning but ineptly configured systems. For example, many consumer products ship with default passwords that are easily discovered in public forums. If a consumer fails to change the login credentials upon installing the device, it is then trivial to seize control of that device for malicious purposes. The relatively unskilled miscreants who knocked the Sony and Microsoft gaming platforms offline over the Christmas holiday in 2014 were using a botnet built from home routers seized in this way (Krebs 2015). Similarly, it is easy for the noninitiate to misconfigure tools, such as using peer-to-peer sharing software in such a way as to inadvertently share far more than was intended (Singer and Friedman 2014: 41–42). Even those who should know better can fail to take basic security precautions, such as neglecting to shield a network with a firewall, failing to set passwords for administrative accounts, or, as was the case with the U.S. Office of Personnel Management, not encrypting sensitive data (Rein 2015).

Application Exploits

More sophisticated attacks exploit weaknesses in software applications to gain control. For example, many websites check user input—such as login credentials—against a stored database. The language generally used to query the database with the user input is called *Structured Query Language*, or SQL (pronounced “sequel”). If care is not taken when writing the code for the website, a user can input a carefully crafted string of characters and have this string interpreted as SQL commands instead of input for a query. Such commands might allow an attacker to see the contents of the database, manipulate or destroy data, or even seize control of the machine hosting the website. This sort of application exploit is called *SQL injection*. Another common mode of attack is the buffer overrun. Here, the attacker provides input that is larger than expected and manages to overwrite space in memory that was reserved for the executing program. This allows the running process to be diverted and for arbitrary code to be executed.

Executive System Exploits

The most complex class of attacks comprises what one might call *executive system exploits*. I intend the term *executive system* to encompass a wide range of physically inhomogeneous

but logically related adaptive control systems. On one end of a spectrum, there is a complex operating system (OS), such as Microsoft Windows, Apple OS X, Unix, Linux, FreeBSD, Android, and iOS that is executed by powerful microprocessors with access to large dynamic memories (like in a laptop computer). Given their extraordinary complexity, any OS is bound to contain weak spots either by design or accident. Of those I just listed, Windows is infamously vulnerable. To give just one example, the Stuxnet worm used no fewer than four ‘zero-day’ exploits against Windows (a zero-day exploit is one which is unknown to the victim and the broader public when it is deployed—they thus have zero days of warning). One of these involved a specially crafted LNK file. A LNK file is a ‘shortcut’ in Windows that points to another file or directory somewhere on the computer. When the user inserts a USB drive into a computer running Windows, the operating system automatically opens a window to browse the USB contents. When the corrupt LNK file is read by the file browser, a flaw in the OS allows malicious code inserted in the file to be executed (LNK Exploits, n.d.; Zetter 2015a).

On the other end of the spectrum is so-called firmware. Firmware is a set of programs or control modules that is somewhere between software, which is easily mutable and intended to be updated during the life of the machine that runs it, and hardware, which is fixed and immutable over the life of the machine. Generally, firmware can be updated—but it seldom is. An enormous range of devices you probably don’t think of as computers do in fact compute and do so under the control of firmware. Examples include TV remotes, elevators, and even toasters. Despite being vastly simpler, the fact that firmware can be changed or that it can process changing data from one or more sensors means that it can be attacked. To give just one example, the firmware that allows a USB flash drive or any other USB device to connect to your laptop currently poses a special risk. In 2014, security researchers announced that they had engineered a way to infect the firmware of the USB controller with self-propagating malicious code (Greenberg 2014a, 2014b; Karsten Nohl and Jakob Lell 2014; Security Research Labs, n.d.).

In between these two extremes is a spectrum of mutable code that more or less directly influences hardware and serves to varying degrees the role of an operating system. So, for instance, many industrial machines and processes—likely including your washing machine—are under the control of a Programmable Logic Controller (PLC). A PLC is a microprocessor-based controller that stores a single, changeable program typically written in a very simple mid-level programming language. These programs generally consist of a series of operations to be executed sequentially or a set of simple logical rules for responding to input from sensors (e.g., when the water level sensor trips, start the agitator). Perhaps the most dramatic cyberattack to target PLCs was that carried out by the digital weapon, Stuxnet. The supersonic centrifuges used to enrich uranium for making nuclear fuel or weapons are controlled by sophisticated (but relatively uncomplicated) PLCs. Stuxnet altered the programs in these PLCs to accomplish an act of extraordinarily subtle sabotage. It caused the supersonic centrifuges to speed up and slow down to wear out their bearings, all the while masking the destruction by playing back false, healthy operating data. In total, the Stuxnet attacks destroyed around 1,000 centrifuges (Zetter 2015a).

Means of Influence

Each of the earlier sections described a vulnerability—a feature of a system that makes it possible to subvert the intended function of that system. But I said little about the means by which those malicious influences could be applied. This provides another dimension along which to classify cyberattacks.

Attacks Through the Communication Network

The internet is exactly what its name implies: an interconnected network of networks. By the early 1970s there were a handful of computer networks, linked collections of computers capable of communicating with one another. ARPANET, a project of the U.S. Advanced Research Projects Agency (ARPA) was the first, connecting research computers located at various U.S. universities. The internet was born when a common protocol for network communication was developed that would allow these disparate networks to communicate with one another, despite their internal differences in hardware (Singer and Friedman 2014: 16–21). The modern internet consists of a collection of physical cables and wires, pieces of hardware such as routers and switches, and the computers they connect. They all communicate with one another using a common protocol. It's a combination of Transmission Control Protocol, which slices a message into 'packets' and handles their reassembly at the intended destination, and Internet Protocol, which concerns the addressing system that directs packets of data passing over the physical network. The pair is often abbreviated TCP/IP. The internet is often conflated with the World Wide Web, which is a collection of resources (webpages, videos, and other data) that: (1) reside on machines connected to the internet, (2) are accessible through a Uniform Record Locator (URL) system of naming (e.g., www.ratiocination.org), and (3) which are heavily interlinked logically by 'hypertext' links (the links you click on in websites). The Web is often a target of cyberattack, and it comprises a lot of the information exchanged on the internet. But cyberattacks are carried out via the internet.

Beyond the internet, there is the Internet of Things (IoT) (Greengard 2015). This is the vast and growing array of objects or things that are not general computing devices like PCs or smartphones but nonetheless have an IP address and communicate with the wider internet. This includes such things as personal fitness wristbands, 'smart' bathroom scales, 'smart' thermostats, webcams, traffic lights, and just about anything else big enough to fit the requisite communications hardware in. It is the sum total of devices that speak TCP/IP. This broader network provides an expanded means of attack.

By far the largest number of existing cyberattack tools are transmitted over the IoT, and the IoT is usually where discussions of cyberattacks and cyberwar terminate. But this, I suggest, is to confuse a general phenomenon with a particular technology. What matters from the perspective of projecting force or influence remotely via computing devices is causal influence between those devices, not the particular communication protocols that make the connections in a chain. There is an even more vast and shifting network of devices beyond the IoT that are connected to each other and, typically, to devices that are in the IoT by an inhomogeneous collection of protocols. I have a small hobbyist computer called a Raspberry Pi. It talks to a microcontroller using a simple protocol known as Serial Peripheral Interconnect (SPI). This microcontroller in turn talks to a tiny temperature sensor that also speaks SPI (despite being the size of a lentil). Other common protocols include I2C, USB, and RS-232. Many of the connections that use these protocols are ephemeral by design. In particular, systems that are considered critical, either for reasons of public safety or protecting sensitive information, are often 'air-gapped.' In other words, they are physically separated (by air) from the network hardware that would allow the systems to be accessed by others via the internet. But air-gapping is never perfect; there is always a way around the physical separation. This is because, in order for systems of any complexity to be useful, they must be updated occasionally with fresh data or software. That means connecting one or more devices that may themselves have been exposed to the broader network. Even though the computers that controlled the Iranian centrifuges at Natanz were air-gapped, Stuxnet was spread to those systems via USB flash drives that were previously

infected by a computer running Windows that was in turn compromised by an attack through the internet. Nothing is out of reach.

Attacks Through Noncommunicative Interaction Chains

Most modern computing devices are ‘embedded’ in larger electro-mechanical systems such as cars, consumer electronics, and military weapon systems. As such, they are intimately bound up with a zoo of sensors and actuators that provide richer ways for machines to project influence on one another, for good or ill, beyond explicit communications protocols. This fact is widely acknowledged with respect to cyber-espionage. The U.S. National Security Agency, for instance, has been aware of (and likely been exploiting) the possibility of surreptitiously reading data from a computer by listening to the electromagnetic radiation generated by the CPU (NSA 1972; Zetter 2014) or video display (Kuhn 2004). Importantly, such unintended channels of influence are also a means of cyberattack. One can, for example, inject malicious signals into a system by exposing its analog sensors to appropriately crafted radio-frequency waveforms (Kune et al. 2013). Alternatively, one can attack by changing the environment around a system’s sensors. To give a concrete example, one might exploit a buffer overrun vulnerability in a microcontroller by manipulating the environment in such a way that a sensor connected to the microcontroller feeds a carefully chosen sequence of unexpectedly large values to the microcontroller. Though this sort of environmental hacking is speculative and, at this point, fairly impractical, this is likely to change the more we saturate the environment with computer-driven actuators and autonomous machines.

Aims of Influence

Finally, it is important to consider the aims—the possible targets—of cyberattacks. Obviously, the specific goals can vary even more widely than the tools used to pursue them. Nonetheless, in the context of nation-state conflict we can discern three kinds of objective.

Psychological

Many, if not most, known cyberattacks aim at psychological disruption. The goals of such an attack include the dissemination of propaganda, intimidation of a particular population, and poisoning the well of information for the public and military planners alike. A notable example took place in Estonia in 2007. That year, this small Baltic nation relocated the Bronze Soldier of Tallinn—a memorial to Russian soldiers who fell in World War II. Russian nationalists took offense and, with the implicit blessing of the Russian government, launched a coordinated DDoS attack against Estonia. Prepackaged tools for conducting a DDoS attack and instructions for using them were disseminated via social media and Russian nationalist websites, and the result was a nation-scale resource attack. It shut down public websites as well as servers running parts of the phone network, the credit-card verification system, and even the nation’s largest bank (Clarke and Knake 2011: 11–16). The effects on computer systems were temporary: the attack lasted three weeks, following which the Estonian internet infrastructure returned to normal. But the psychological effect was clearly more profound (Singer and Friedman 2014: 110–111). The Estonian foreign minister asserted that, “The attacks are virtual, psychological and real” (Urmas Paet 2007). Making a case before fellow NATO members that the rules of their alliance demand a joint response, the Estonian prime minister likened the cyberattack to a naval blockade (Rid 2013: 7).

Tactical

On a quiet night in September of 2007, a squadron of Israeli jets leveled a large building complex located well within Syria's borders. The complex was an illicit nuclear weapons facility designed and built with the assistance of North Korea. Despite sinking billions of dollars into electronic air defense systems, the Syrians never saw the Israeli planes coming or going. Ultimately, it was revealed that Israel had used a cyberattack to blind the Syrian radar. Precisely how they did this remains uncertain; possibilities include hacking the radar directly by feeding it a carefully crafted return signal from a stealth Unmanned Aerial Vehicle (UAV), or the incorporation of a malicious code by an Israeli agent into the software that ran the Russian-built systems (Adee 2008; Clarke and Knake 2011: 1–8; Follath and Stark 2009; Singer and Friedman 2014: 126–128). Whatever the mechanism used in 'Operation Orchard,' as it was called, it is an excellent illustration of the tactical aims to which cyberattacks can be put. Another is the coordinated use of cyberattacks in support of the Russian air, land, and sea assault on Georgia during the 2008 dispute over South Ossetia. A combination of cyberattacks, including DDoS and SQL injection, were closely coordinated with the use of conventional forces to shut down Georgian government websites in order to cut off communication with the Georgian people and foreign governments (Carr 2011: 3; Singer and Friedman 2014: 125). More recently, Iran captured a U.S. RQ-170 stealth drone, apparently by spoofing the GPS signal it uses to navigate (Johnson et al., n.d.; Peterson and Faramarzi 2011; Rawnsley 2011). Such attacks—by disabling or commandeering enemy equipment, clouding communications on the battlefield, or misdirecting enemy attention—are designed to play key roles in small-scale, short-term military maneuvers.

Strategic

Finally, as suggested by the scenario I sketched at the outset of this chapter, cyberattacks can be used strategically to degrade the infrastructure, economy, and war-fighting capability of an adversary. Stuxnet was aimed at slowing the Iranian nuclear program. Though its actual impact is debatable, it's certain that the worm could have been far more destructive if it hadn't been designed for stealth. It had the capability to destroy the centrifuges at Natanz all at once. Similar strategic attacks against a nation's warfighting capabilities are both possible and doubtless under active development. We are also likely to see cyberattacks aimed at general civilian infrastructure with military implications, such as the power grid, manufacturing facilities, and communications. For example, in January of 2015, hackers seriously damaged a German steel mill by preventing a controlled shutdown of a blast furnace (Zetter 2015b). In 2008, the U.S. Federal Bureau of Investigation reported that counterfeit routers, switches, and interface cards—key pieces of equipment for directing traffic on computer networks—imported from China posed a risk to American networks (Krigsman 2008). Specifically, a cyberattack could be launched using a 'back door' in the code controlling the rogue equipment that would allow an attacker (presumably China) to shut down large swathes of the U.S. communications network. While there is some controversy over the possibility of achieving lasting strategic effects with a cyberweapon (see, e.g., Lewis 2011), it is clear that much of the critical war-fighting infrastructure of a wired country like the United States is at least temporarily vulnerable.

What Makes an Action a Cyberattack?

Having briefly surveyed some examples of cyberattacks, we are now in a position to look for uniting features. What makes an activity a cyberattack? What, specifically, what makes

it ‘cyber’? What, if anything, separates such an attack from any other act of violence against property or persons?

In light of the earlier examples, we can reject two of the most common proposals. To begin with, some (e.g., Shakarian et al. 2013: 2) speak of cyberwarfare as warfare that takes place in a special domain of combat called *cyberspace*, just as the ocean is the special domain of naval warfare. But where exactly is ‘cyberspace’? It’s variably identified with “the information environment” (“DOD Dictionary of Military and Associated Terms,” n.d.), “inter-connected networks of information technology infrastructures” (Tabansky 2011: 78), or merely a “notional environment” (“cyberspace, n.,” 2015). It isn’t a literal spatial volume in which you can walk around. But the centrifuges that Stuxnet destroyed were most definitely to be found in real, physical space, a space not consisting of information technology infrastructures.

Others have characterized cyberattacks in terms of their intended target. The National Research Council, for example, defines cyberattacks as “deliberate actions to alter, disrupt, deceive, degrade, or destroy computer systems or networks or the information and/or programs resident in or transiting these systems or networks” (Owens et al. 2009: 1). This, too is somewhat unsatisfying. While it is true that the aim of a DDoS attack is to cripple a piece of information technology infrastructure (e.g., someone’s webpage, or, more seriously, the network used by the military to coordinate remote actions) that is not obviously the case in general. For instance, when the Iranians downed the RQ-170 drone, they were aiming to influence a piece of physical hardware, not disrupt information. In my introductory scenario, the targets were weapons manufacturing plants, not information infrastructure. This approach thus seems to confuse the means with the aims of cyberattack.

A final and more promising approach views cyberattacks as a special mode of attack. For example, Singer and Friedman say that cyberattacks “use digital means, a computer action of some sort. . . . Instead of causing direct physical damage, a cyberattack always first targets another computer and the information within it” (2014: 68–69). Clarke and Knake define “cyber war” as “actions by a nation-state to penetrate another nation’s computers or networks for the purposes of causing damage or disruption” (2011: 6). And the Tallinn Manual, an internationally crafted, NATO-sponsored document often cited as an authoritative source on cyber conflict defines a cyberattack as a “cyber-operation . . . reasonably expected to cause injury or death to persons or destruction to objects” (Schmitt 2013). These explications are close to capturing all of the actual known examples of cyberattacks, and at the same time clearly delineating what sets them apart from conventional kinetic attacks. Like a kinetic attack, the targets and venues of cyberattacks are wide and varied. But *unlike* a kinetic attack, cyberattacks are essentially indirect—they are mediated by one or more computers, things that are not inherently weapons. Taking care not to obscure an underlying feature by imposing too much detail that is only an accidental feature of existing technology, we can refine the proposal as follows.

Cyberattack: an intentional effort to apply force against an adversary where the intended effect is the terminus of a causal chain in which at least one link crucially involves a computation that: (1) is specified or determined in advance by the attacker; and (2) induces behavior contrary to that intended by the designers of the device carrying out the computation.

The term *force* in the definition requires some clarification. As Stone (Stone 2013) points out, it is important to distinguish three concepts from one another: force, violence, and lethality. All of these have something to do with the notion of warfare. Force, as the great theoretician of war, Carl von Clausewitz (1993), understood it, is literally that which effects or compels physical change. This is more general than either violence, which involves damage or destruction

of persons or things, or lethality, which involves the killing of people. Force does not entail violence, nor does violence entail lethality. Though war overall generally involves lethality, many commonly recognized conventional attacks or acts of war involve only violence or force. For instance, bombing an empty bridge is violent but not lethal. Jamming a radar station with a radio beam is neither lethal nor destructive. Yet both are generally considered attacks. While relatively few are potentially lethal or even violent, all of the cyberattacks described earlier involve force.

Cyberwarfare

Recall that my strategy for clarifying the notion of cyberwarfare was to explicate the more straightforward notion of cyberattack and then to understand cyberwarfare as warfare—whatever exactly that may be—involving one or more cyberattacks.

Cyberwarfare: warfare involving one or more attacks in which there is an intentional effort to apply force against an adversary, and where the intended effect is the terminus of a causal chain in which at least one link crucially involves a computation that is specified or determined in advance by the attacker and that runs contrary to the intended operation of the computing system.

With this more or less precise definition in mind, we can identify a number of special features of cyberwarfare. Perhaps the most widely discussed of these is the *difficulty of attribution*. Cyberattacks of all sorts are notoriously difficult to attribute to any particular actor. Simple attacks like DDoS can be made through botnets whose constituent machines are scattered around the world in innocent countries, making it difficult if not impossible to determine the source. For more sophisticated attacks, it's possible for the attackers to leave clues behind, such as names or other information buried in compiled code, but there is no way to distinguish a genuine clue inadvertently left behind from a malicious plant that incriminates the wrong nation-state. Practically speaking, this uncertainty can be exploited to produce mayhem of all sorts. In 2007, Israel conducted a military simulation that ended with the United States on the brink of conflict with Israel and the Israelis preparing to invade Syria and Lebanon, largely on the basis of misattributed cyberattacks that were actually (within the simulation) launched by Iran (Zetter 2015a: 379–380).

Another characteristic feature of cyberwarfare vis-a-vis traditional kinetic warfare is *asymmetry*. Actually, there are two sorts of asymmetry to worry about. The first, and most widely cited, concerns the low cost of producing and deploying a cyberweapon relative to a conventional weapon. That cost is so low that anyone in the world with access to a computer and an internet connection can launch a cyberattack. This has led to some hyperbolic hand-wringing over a lone teenage hacker bringing a nation-state to its knees, but this concern is based on an oversimplified view of the nature of cyberattacks. As we saw earlier, these range widely in sophistication, destructive power, and strategic value: the greater the power and precision of a cyberattack, the greater the resources required to pull it off. On the one hand, it's true that a lone hacker can launch a very successful DDoS attack, but shutting down webpages is unlikely to cripple the warfighting capability of a modern nation. On the other hand, attacks like Stuxnet have substantial strategic value, but they require enormous infrastructure and talent to produce. Stuxnet, for instance, was built by multiple teams of highly trained and talented programmers and required detailed knowledge (and almost certainly, physical copies) of the hardware being used in the nuclear facility at Natanz. Only a nation-state is likely to be able to assemble such resources.

There is, however, a second and more worrisome asymmetry in vulnerability. It is the case that even isolated, otherwise technologically stunted nations like North Korea have the resources to build extensive cyberwarfare capabilities (Carr 2011: 246–247). In fact, it's in principle possible for these capabilities to rival those of a large, wealthy, technologically advanced nation like the United States. But a nation like North Korea—with virtually no infrastructure to speak of—is not itself vulnerable to cyberattack. While an effective cyberattack on the U.S. electrical grid could result in significant loss of money, life, and warfighting capacity, destroying the electrical infrastructure of North Korea would mean relatively little. This means that cyberwarfare offers a new and dangerous kind of warfighting asymmetry between nation-states, one for which there can be no analog of the 'mutually assured destruction' that kept nuclear arms from being deployed.

Finally, cyberwarfare is, at least in principle, subject to greater automation than conventional warfare. As many are fond of pointing out, a cyberattack travels at the speed of electrical impulses. Targets distributed throughout the world can be hit nearly instantaneously. With little time to react, some—such as the U.S. NSA—have thought it prudent to build systems to retaliate or to attack automatically when trigger conditions are met (Gillum 2014). There are at least two sorts of automation to consider. First, systems can be deployed to retaliate or attack automatically when certain predetermined conditions are met. If more than one nation-state employs such a system, there is the unprecedented risk of a chain reaction of increasingly dire automatic retaliation from both sides in a dispute (see (Danks and Danks 2013) for an overview of this possibility and its moral implications). Second, one can also imagine more adaptive cyberweapons that are endowed with the capacity to adjust plans and learn from experience in order to reach their targets. This sort of weaponized machine learning algorithm raises a slew of worries about doomsday scenarios (e.g., Barrat 2013).

An Inevitable Trend

As I indicated at the outset, the aim of this chapter is to get a handle on just what cyberwarfare is so that we can begin to answer important questions about its nature and morality. The upshot of this effort is a view of cyberwarfare as warfare incorporating attacks that are constituted by causal chains involving computation in at least one critical step. I've tried to frame this conception independent of the details of current technology, details that change fast and seem superfluous to the phenomenon in question. While my proposal leaves some key ideas vague, it highlights a very robust trend in the recent history of technology. It is widely known that computing power has grown ever cheaper. What is not often discussed is the extent to which causal chains involving computation continue to lengthen and ramify. Devices with a spectrum of computing power permeate the environment and connect ever more physical processes by computational causal chains. In the late 1960s and early 1970s we began to computerize our industrial processes. With the birth of the internet in the 1990s, distant control systems were linked together: however indirectly, a PLC in an automobile factory could influence the temperature of a blast furnace at a smelting plant. With the 'sensor revolution' of the early twenty-first century, computational measurement and control spilled out into the broader world of consumer products and private dwellings. Now the refrigerator in my neighbor's house is connected (albeit by a very lengthy chain) to that same blast furnace. All signs suggest that this trend will continue, and more sensors and cheaper computing power mean that great swathes of the built environment (and also, perhaps, the natural environment) will fall under the sway of long causal chains that depend on computation. By extension, the turf of cyberwarfare—the ways in which cyberattacks can impact the infrastructure, environment, and daily life of an

adversary—will only grow. The future will offer modes of coercion utterly alien to conventional armed conflict.

I've suggested that the phenomenon of cyberattack, and by extension cyberwarfare, is genuinely novel. In a sense, a cyberattack is a metatool, a technology for manipulating technologies that in turn manipulate the physical world. This crude understanding of the nascent phenomenon of cyberwarfare raises a host of questions. Some concern the technology itself. What does this capacity for indirect influence and coercion mean for the development of international conflict? Given the increasing complexity of cyberattacks and the systems they exploit, there is pressure to develop autonomous responses. But is it even possible to devise effective automated systems for cyberdefense or attack that learn fast enough and respond appropriately? There are hints already of an extraordinary arms race under way, one that requires ever more sophisticated methods for generating automated military software, perhaps to the point where those methods of generation themselves will have to be automated. Will such weapons constitute human artifacts or something else? Other questions concern the nature of warfare: will cyberwarfare increase or decrease violence in international conflict? Will it make war more or less likely? There is a case to be made on either side (Clarke 2009; Rid 2013). Finally, and perhaps most pressingly, cyberwarfare raises many new ethical issues. Because of the difficulty of attribution, one must worry whether it is ever possible to satisfy the epistemic demands of the just use of force. If we can never be reasonably sure who attacked us, how can any retaliation be just (Eberle 2013)? Assuming retaliation can be justified, what about retaliation carried out by machines without human intervention (Danks and Danks 2013)? Is the dominant moral framework of Just War Theory even applicable to these novel modes of violent and nonviolent coercion (Bringsjord and Licato 2015)? These questions only scratch the surface of what we and the next generation will want to know about cyberwarfare and the curious confluence of technologies that make it possible.

References

- Adee, S. (2008). The Hunt for the Kill Switch. *IEEE Spectrum*, 45, 34–39. doi:10.1109/MSPEC.2008.4505310
- Barrat, J. (2013). *Our Final Invention: Artificial Intelligence and the End of the Human Era*, 9.1.2013 edition. New York: Thomas Dunne Books.
- Bringsjord, S., and Licato, J. (2015). By Disanalogy, Cyberwarfare Is Utterly New. *Philosophy & Technology*, 28, 339–358.
- Carr, J. (2011). *Inside Cyber Warfare*, 2nd edition. Sebastopol, CA: O'Reilly Media, Inc.
- Clarke, R. (2009). War From Cyberspace. *The National Interest*, (104), 31–36.
- Clarke, R. A., and Knake, R. (2011). *Cyber War: The Next Threat to National Security and What to Do About It*, Reprint edition. New York: Ecco.
- Clausewitz, C. von (1993). *On War, Everyman's Library*. New York: Knopf.
- Clock Ticking on Worm Attack Code (2009). *BBC*.
- Cyberspace (2015). OED Online.
- Danks, D., and Danks, J. H. (2013). The Moral Permissibility of Automated Responses During Cyberwarfare. *Journal of Military Ethics*, 12, 18–33. doi:10.1080/15027570.2013.782637
- DOD Dictionary of Military and Associated Terms (n.d.). Available at: www.dtic.mil/doctrine/dod_dictionary/data/c/10160.html [Accessed October 26, 2015].
- Eberle, C. J. (2013). Just War and Cyberwar. *Journal of Military Ethics*, 12, 54–67.
- Follath, E., and Stark, H. (2009). The Story of “Operation Orchard”: How Israel Destroyed Syria's Al Kibar Nuclear Reactor. *Spieg*. Online.
- Friedman, J. 1972. Tempest: A Signal Problem. *NSA Cryptologic Spectrum*, 4, 26–30.

- Giannetos, T., Dimitriou, T., Krontiris, I., and Prasad, N. R. (2010). Arbitrary Code Injection Through Self-Propagating Worms in Von Neumann Architecture Devices. *The Computer Journal*, 53, 1576–1593.
- Gillum, J. (2014). NSA Planned Automated Cyberwarfare Program. *The Christian Science Monitor*.
- Goodin, D. (2016). First Known Hacker-Caused Power Outage Signals Troubling Escalation. *Ars Tech*. Available at: <http://arstechnica.com/security/2016/01/first-known-hacker-caused-power-outage-signals-troubling-escalation/> [accessed January 7, 2016].
- Greenberg, A. (2014a). The Unpatchable Malware That Infects USBs Is Now on the Loose. *WIRED*. Available at: www.wired.com/2014/10/code-published-for-unfixable-usb-attack/ [Accessed October 3, 2015].
- Greenberg, A. (2014b). Why the Security of USB Is Fundamentally Broken. *WIRED*. Available at: www.wired.com/2014/07/usb-security/ [Accessed October 4, 2015].
- Greengard, S. (2015). *Internet of Things*. Cambridge, MA: MIT Press.
- Johnson, R. (December 5, 2011, 28, 849, 25, n.d.) Meet The Russian Avtobaza—Iran’s Possible Drone Killer. *Business Insider*. Available at: www.businessinsider.com/meet-the-russian-avtobaza-irans-possible-drone-killer-2011-12 [Accessed September 9, 2015].
- Krebs, B. (2015). Lizard Stresser Runs on Hacked Home Routers—Krebs on Security. *Krebs Security*.
- Krigsman, M. (2008). FBI: Counterfeit Cisco Routers Risk “IT Subversion”. *ZDNet*. Available at: www.zdnet.com/article/fbi-counterfeit-cisco-routers-risk-it-subversion/ [Accessed October 9, 2015].
- Kuhn, M. G. (2004). Electromagnetic Eavesdropping Risks of Flat-Panel Displays. In D. Martin and A. Serjantov (eds.) *Privacy Enhancing Technologies, Lecture Notes in Computer Science*. Berlin Heidelberg: Springer, 88–107.
- Kune, D. F., Backes, J., Clark, S. S., Kramer, D., Reynolds, M., Fu, K., Kim, Y., and Xu, W. (2013). Ghost Talk: Mitigating EMI Signal Injection Attacks Against Analog Sensors. In 2013 IEEE Symposium on Security and Privacy (SP). Presented at the 2013 IEEE Symposium on Security and Privacy (SP), 145–159. doi:10.1109/SP.2013.20
- Lewis, J. A. (2011). Cyberwar Thresholds and Effects. *IEEE Security and Privacy Magazine*, 9, 23–29. doi:10.1109/MSP.2011.25
- LNK Exploits (n.d.). *Lavasoft*. Available at: www.lavasoft.com/mylavasoft/securitycenter/whitepapers/lnk-exploits [Accessed September 29, 2015].
- Meserve, J. (2007a). Sources: Staged Cyber Attack Reveals Vulnerability in Power Grid. *CNN*. Available at: www.cnn.com/2009/us/26/power.at.risk/index.html [Accessed January 11, 2016].
- Meserve, J. (2007b). Staged Cyber Attack Reveals Vulnerability in Power Grid. *CNN*. Available at: www.cnn.com/2007/US/09/26/power.at.risk/ [Accessed January 11, 2016].
- Naone, E. (2009). Meters for the Smart Grid. *MIT Technology Review*. Available at: www.technologyreview.com/hack/414820/meters-for-the-smart-grid/ [Accessed January 11, 2016].
- Nohl, K., and Lell, J. (2014). BadUSB—On Accessories that Turn Evil. *Black Hat*.
- Krebs, B. (2015). Lizard Stresser Runs on Hacked Home Routers. *Krebs on Security*. Krebs Secur. Available at: <https://krebsonsecurity.com/2015/01/lizard-stresser-runs-on-hacked-home-routers/>
- Krigsman, M. (2008). FBI: Counterfeit Cisco Routers Risk “IT Subversion”. *ZDNet*. Available at: www.zdnet.com/article/fbi-counterfeit-cisco-routers-risk-it-subversion/ [Accessed October 9, 2015].
- Kuhn, M. G. (2004). Electromagnetic Eavesdropping Risks of Flat-Panel Displays. In D. Martin and A. Serjantov (eds.) *Privacy Enhancing Technologies, Lecture Notes in Computer Science*. Berlin and Heidelberg: Springer, 88–107.
- Kune, D. F., Backes, J., Clark, S. S., Kramer, D., Reynolds, M., Fu, K., Kim, Y., and Xu, W. (2013). Ghost Talk: Mitigating EMI Signal Injection Attacks against Analog Sensors. In 2013 IEEE Symposium on Security and Privacy, 145–159. doi:10.1109/SP.2013.20
- Lewis, J. A. (2011). Cyberwar Thresholds and Effects. *IEEE Security and Privacy* 9, 23–29. doi:10.1109/MSP.2011.25
- LNK Exploits. (n.d.) *Lavasoft*. Available at: www.lavasoft.com/mylavasoft/securitycenter/whitepapers/lnk-exploits [Accessed September 29, 2015]
- Meserve, J. (2007a). Sources: Staged cyber attack reveals vulnerability in power grid. *CNN*. URL www.cnn.com/2007/USNSA.

- Owens, W. A., Dam, K. W., and Lin, H. S. (eds.) (2009). *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities*. Washington, DC: National Academies Press.
- Peterson, S., and Faramarzi, P. (2011). Exclusive: Iran Hijacked U.S. Drone, Says Iranian Engineer (Video). *The Christian Science Monitor*.
- Rawnsley, A. (2011). Iran's Alleged Drone Hack: Tough, But Possible. *WIRED*. Available at: www.wired.com/2011/12/iran-drone-hack-gps/ [Accessed September 20, 2015].
- Rein, L. (2015). Top House Republican Calls on OPM Director to Resign Over Employee Data Breach. *Washington Post*. Available at: www.washingtonpost.com/news/federal-eye/wp/2015/06/16/lawmaker-tells-opm-chief-you-failed-utterly-and-totally-to-prevent-data-breach-by-chinese/?utm_term=.1ceacef64c05 [Accessed September 20, 2015].
- Rid, T. (2013). *Cyber War Will Not Take Place*. New York: Oxford University Press.
- Schmitt, M. N. (ed.) (2013). *Tallinn Manual on the International Law Applicable to Cyber Warfare*. Cambridge: Cambridge University Press.
- Security Research Labs (n.d.) *Turning USB Peripherals Into BadUSB*.
- Shakarian, P., Shakarian, J., and Ruef, A. (2013). *Introduction to Cyber-Warfare: A Multidisciplinary Approach*, 1st edition. Amsterdam; Boston, MA: Syngress.
- Singer, P. W., and Friedman, A. (2014). *Cybersecurity and Cyberwar: What Everyone Needs to Know*, 1st edition. Oxford ; New York: Oxford University Press.
- Stone, J. (2013). Cyber War Will Take Place! *Journal of Strategies Studies*, 36, 101–108. doi:10.1080/01402390.2012.730485
- Tabansky, L. (2011). Basic Concepts in Cyber Warfare. *Military and Strategic Affairs*, 3, 75–92.
- United States Environmental Protection Agency (2001). Chemical Accidents From Electrical Power Outages (Chemical Safety Alert No. EPA 550-F-01–010).
- Urmas Paet (2007). Statement by the Foreign Minister Urmas Paet. *Eesti Päeval*. Available at: <http://epl.delfi.ee/news/eesti/statement-by-the-foreign-minister-urmas-paet?id=51085399> [Accessed December 23, 2015].
- Zetter, K. (2014). How Attackers Can Use Radio Signals and Mobile Phones to Steal Protected Data. *WIRED*. Available at: www.wired.com/2014/11/airhopper-hack/ [Accessed November 13, 2015].
- Zetter, K. (2015a). *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*. New York: Broadway Books.
- Zetter, K. (2015b). A Cyberattack Has Caused Confirmed Physical Damage for the Second Time Ever. *WIRED*. Available at: www.wired.com/2015/01/german-steel-mill-hack-destruction/ [Accessed September 23, 2015].

The Cloud

Craig A. Condella and Julie C. Swierczek

Strepsiades: Zeus! Socrates, you must tell me, who are these ladies singing this amazing song? Are they some new breed of female idols?

Socrates: No, no, no. They are the heavenly Clouds, magnificent goddesses for men of leisure. They grace us with our intellect, argumentative skills, perception, hyperbolization, circumlocution, pulverization, and predomination.

(Aristophanes 2000: 23–24)

Socrates famously proclaims philosophy to be the “practice for dying and death” (Plato 1997: 55). Such an odd description of philosophy points to a basic metaphysical assumption, namely that the really real—that which truly is—is unchanging and incorruptible. While his friends lament Socrates’s impending drinking of hemlock, Socrates himself views his death as a release from an embodied existence, one where knowledge is imperfect and fleeting. True knowledge and, by extension, the fulfillment of the philosopher, can only be had in a realm that is timeless and unchanging, a nonphysical world of pure ideas, or what philosophers since have come to understand as the Platonic forms. As such, Socrates’s death sentence is a cause for celebration rather than lamentation.

The comedic playwright Aristophanes, a contemporary of Socrates, is perhaps the first to lampoon this notion of an otherworldly realm, his *Clouds* makes light of the perceived idleness of philosophers and their presumably ill-begotten association of the true with the ethereal. Such early satire, however, did little in deterring subsequent thinkers in the Western philosophical tradition from privileging the transcendent, pure, nonbodily, and unchanging over the immanent, corruptible, embodied, and transient. From Dante’s ascent through the perfection of an Aristotelian-inspired heaven to Descartes’s identification of his own essence as mind rather than body to the Hegelian realization of Absolute Spirit, Western philosophy evinces many attempts to overcome the messiness of our everyday reality in the hope of discovering that which is eternal and true. Making much of this tendency in the thinkers who preceded him, the twentieth-century German philosopher Martin Heidegger maintains that the Western philosophical tradition as a whole is a Metaphysics of Presence wherein philosophers have tried to *fix* Being and thereby render *permanently present* that which otherwise seems to be in a state of perpetual change (see especially Heidegger’s *Introduction to Metaphysics*). Like Aristophanes many centuries before him, Heidegger looks askance at such attempts; nevertheless, he believes they reveal much in the way of explaining the long history of what has culminated in the modern technological age.

In this light, the Cloud represents far more than a technological achievement. Indeed, it is more like the realization of a dream that reason first conjured in Athens in the fifth century BCE. In the popular imagination, the Cloud allows us to overcome that which Socrates would have taken to be insurmountable—at least this side of death—namely, the limitations of space and

time. In terms of *space*, the term itself suggests a place more virtual than real, one that allows for the storage and sharing of information without regard for the confines of the physical world. To think in terms of a few examples, the Cloud puts to shame the limits on document, photo, or music storage that our filing cabinet, hope chest, and CD rack had heretofore imposed. And yet, the ability to store so much more information is only half the story, since the nature of this storage is fundamentally different from what was previously thought possible. When we upload something to the Cloud, it seems to immediately take on a timeless quality, being now immune to the ravages of time. No longer need we concern ourselves with the withering of documents, the fading of photographs, or the ever-accumulating scratches on CDs, for the Cloud allows us to not only store but preserve in a way that Socrates could not have imagined. Finally, born out of web-based communication, it should not go unnoticed that the seemingly infinite storage and flawless preservation that the Cloud provides extends well beyond private use; indeed, the Cloud enables the sharing of information with family, friends, collaborators, and even strangers in a way that makes someone half a world away as close—or in some ways even closer—than the person in the next room.

As with any technology, however, what the Cloud promises must be tempered by its real limitations and darker, oftentimes ignored, possibilities. Perhaps the Cloud does make manifest what philosophers of previous ages could only have dreamt of, but this does not eliminate the possibility that such dreams, if unchecked, can turn into nightmares. Our awareness of both the promise and perils of the Cloud becomes crucial if we are to retain at least some level of control, even while acknowledging the impossibility—and perhaps undesirability—of complete mastery.

The Cloud We Imagine

Thinkers since the middle of the twentieth century have variously christened our present age (pervaded as it is by the wonders and fears that accompany rapid technological growth) as everything from the digital age to the nuclear era to the age of cybernetics. Whichever title we choose, one feature of contemporary society remains undeniable—the explosion of information that modern technology has engendered. Writing before, though perhaps anticipating, the possibilities of the 2.0 version of the World Wide Web, Neil Postman, in his 1992 work *Technopoly*, speaks to the ways in which importantly new technologies can have ecological impacts on society, not just adding a new dimension to the way we live, but changing our world entirely. This is especially true when such change involves the generation and dissemination of information previously limited to a relatively small number of people. The classic example here is Johannes Gutenberg's printing press, a technology as responsible as any other for delivering Europe into the modern era. Though it would be hard to deny the positive powers of the printing press—leading as it did to an increase in books, readers, and ultimately knowledge—it is similarly difficult to deny that the printing press led to something of an information crisis that involved everything from the ideological battles of the Protestant Reformation to the more practical question of how to teach the suddenly necessary skills of reading and writing to the largely illiterate working class. In short, Europe needed time to adjust to the exponential growth in information that the printing press brought about, adjustments that involved book format standardization, more precise methods of cataloging, and the birth of the modern school system.

More recently, while twentieth-century technologies such as the radio and television provided new and faster ways of communicating ideas to increasingly larger audiences, the digitization of information and the birth of the internet have arguably created the greatest information crisis since Gutenberg's fifteenth-century invention. In this context, the functions of the Cloud become incredibly important.

A simple example here may prove helpful in showing how the Cloud helps the average user deal with what Postman describes as the information glut that has become a natural, though

regrettable, feature of our modern technological society (Postman 1992: 56–70). Consider picture taking. Similar in some ways to the more traditional (and increasingly rare) practice of film photography, digital photography is in many ways a fundamentally different experience. If for no other reason than the cost of film and its subsequent development, film photographers tend to take pictures more judiciously than their digital counterparts. As people who grew up using film cameras, we can recall going to sporting events and taking five to ten photographs in the hope that two or three of them might turn out well. Now, with digital cameras, there is seemingly no need for such discretion. We can take as many photographs as we want, delete or even edit them on the spot, and otherwise just store them all on a computer. As a result of this move from analog to digital, the increase in the number of photographs that many people take has been more exponential than additive, the fact that so many now carry cameras in their pockets (as part of their phones) playing a crucial role in this increase as well.

But where do we now store all of these photos? On the surface of it, this does not immediately present a tremendous challenge, as the shoeboxes and linen closets of old are simply replaced by some space on my hard drive. Over time, however, hard drive storage alone became unsatisfactory. First, hard drive storage is limited, and though those limits are ever-increasing, so too are the demands for space. Beyond these limits, however, are the oftentimes disappointingly short lives of our hard drives, a feature that anyone who has lost treasured photos because of a hard drive crash can well attest to. So, while digital photos are immune to the water or heat damage that have so often plagued photos of the analog variety, their dependence on a hard drive revealed an Achilles heel of their own. Hence the need for external hard drives and the need to ‘back up’ our information from time to time, the digital equivalent of defrosting your freezer. All of this, from limited hard drive space to the need for external drives to the self-imposed reminders to back up my data, seem to still call for something more, namely a more permanent way to store more photos, a better way to organize them, and—perhaps most importantly—a way to share them with others. Here enter the possibilities afforded to us by the Cloud. Now we can upload my photographs (and any other digitized information for that matter) and render them accessible to others, all in a way that is not susceptible to the slings and arrows of outrageous fortune or—what amounts to the same—the stability of the hard drive in my laptop. In short, the Cloud allows me to store, organize, and share information. As such, it acts as something of a meta-tool that allows individuals to negotiate current the information crisis that the rest of our tools have done so well to create.

Viewed in its most positive light, the Cloud not only helps us to deal with the digital universe, but to flourish within it. Beyond its ability to store things like photos, music, and legal documents (to name but a few), the Cloud allows us to research and collaborate in ways that were previously impossible. To take this very chapter as but one example, the use of Google Docs allowed its authors to work together without really being together at all. Whereas co-authorship in the past would require the real proximity of the parties involved, the Cloud once again allows us to ignore the limitations of space and time, providing a *virtual* forum for the exchange of *real* ideas. Were Socrates alive today, we might imagine him embracing the transcendent possibilities of the Cloud as something of a Platonic heaven right here on Earth.

The Cloud As It Is

Of course, what we *imagine* the Cloud to be is one thing. What it *actually is* is another. The U.S. National Institute of Standards and Technology enumerates the core characteristics of the Cloud, which can be paraphrased as services

available when required (‘on demand’), without the need for lengthy procurement and configuration processes; available on standard networks such as the internet, without special

requirements for obscure or proprietary networking, protocols, or hardware; able to offer additional capacity as demand increases, and less as demand falls ('elastic'); [and] capable of only billing customers for the storage they use.

(Beagrie, Charlesworth and Miller 2014)

These are key features common to all tiers of cloud computing.

When most consumers think of the Cloud, they are thinking of one tier of cloud computing; however, there are three main tiers. Imagine Alice wants to sell her vast array of widgets online. To do that, she needs to invest in her information technology (IT) infrastructure. She needs servers to run her website, databases to serve and contain her product information, backup storage for all of her data, a network connection that can handle enough traffic for her business, and tools to help balance web traffic and transactions on her site. She also needs a data center for storing and securing all of the computing and network equipment, plus preventive mechanisms to reduce risk to her equipment in the event of fire or other catastrophic events. This adds up to a significant investment in equipment, facilities, and personnel, and the costs may be prohibitive for a new business.

However, Alice doesn't have to purchase all of those resources outright: she can turn to the Cloud. The first tier of cloud computing is IaaS, Infrastructure as a Service. In this model, a cloud computing provider buys all of the equipment and network connectivity and sets up the secure data center where it will be housed. The provider then pools these resources and offers to sell access to the computing and network power of them, much like the electric company sells electricity without consumers having to own pieces of the electrical grid. Alice is an IT professional, so she will take care of installing operating systems, configuring her network, running her website, scheduling her backups, patching her operating system, and performing other administrative tasks. The benefit to Alice is that she does not have to make a significant initial investment in these resources; rather, she pays only for the amount of computing resources that she needs.

This relationship is beneficial to both Alice and the cloud computing provider. Imagine Alice starts her business requiring two units of computing power. (These units are an abstraction.) Perhaps the most economical servers she could buy hold eight units of computing power. Until Alice's business becomes successful, she would have six units of computing power unused, but she would still have to pay for all of the overhead required to run the full eight units of computing power.

When the cloud services provider owns the infrastructure, it can sell access for just two units to Alice, and sell access to the remaining units to other customers. Additionally, if Alice decides she needs to increase her computing power, she can contract with the provider to do so. This is particularly useful for businesses that need different amounts of computing power at different times of the year, such as the December holidays, or even different times of the day, like a video streaming service having more customers in the evening than at any other time of the day. Customers can contract with the provider to have access to more resources at certain times; when the demand reduces, those resources can be repurposed for other uses. Cloud service providers often provide resources at lower cost during off-peak hours, and there are significant savings if a user can be flexible about availability. For example, if some scientists need to run massive calculations in the next month, but the specific date and time of those calculations is unimportant, then the cloud service provider can fit them in between peak uses of the servers by their higher-paying business customers.

The next level of cloud computing is PaaS, or Platform as a Service. This model is a step beyond the IaaS model. In PaaS, the customer pays for everything in the IaaS model, plus a little more for some additional IT support: the provider installs and maintains the operating system on the computers and takes care of the network management as well. So, if Bob wants to start a business and he doesn't have time to do all of the IT functions himself, he can pay the

provider to do some of them for him. Bob can still run his own software programs as he sees fit, but he pays the provider to install operating system patches, to manage his network security, to ensure that his backups are being run appropriately, and to do other administrative tasks.

There are other services typically included in PaaS. For example, when Alice adds more computing power to her system, she has to do a number of tasks. She has to change settings in her system to balance the load across all of her servers. She has to update her network connections so that her customers can access the additional servers. In the PaaS model, the provider typically takes care of those tasks on the customer's behalf. Bob can enter into an agreement where he can set conditions so that if x many people try to access his servers at once, the vendor will automatically add x numbers of servers to handle the traffic, and the vendor will handle the load balancing and other related tasks. Bob doesn't have to configure the equipment himself the way that Alice does. There are several well-known PaaS providers, such as Amazon Web Services, Rackspace, or Microsoft Azure.

The next layer in cloud computing is SaaS, Software as a Service. This is what most consumers mean when they talk about the Cloud. Under this model, the consumer pays for access to a particular software program, typically with a storage component, running on a platform on infrastructure owned by someone else. So, if Carol wants to store her family photos online so that she can share them with other family members, she can sign up for a cloud storage service intended for use by individual consumers. It may be odd to think of cloud storage as software, but most cloud storage used by consumers is not just storage of files on a hard drive. There are usually added features that are operated by software, such as the ability to share files or folders with other through links or via email. Or there may be a desktop application that syncs files on the desktop to the Cloud and across other devices with the syncing application installed. Another SaaS product many consumers use is online music library systems, where they can upload and manage their music online, and have access to it anywhere. Many mobile device applications are SaaS. One of the more well-known SaaS products is web-based email; the email software and the email itself is stored on equipment owned by someone other than the end user. (Note that we used to call this model *web-based* before *Cloud* became a popular term.) Probably the most ubiquitous SaaS products of all are social networks such as Facebook.

Another way that people think about the Cloud is as any computing activity where the consumer is not relying on software and/or storage located on the consumer's own hardware. The consumer has to rely on a service provider to either run the software, provide storage, or both. Google Chromebooks are built on this idea; very little software is run on the computer itself. Much of the computing that takes place happens through the browser, using online applications. Most of the content produced by the consumer is also stored in the Cloud.

A Cloud of Concerns

Since most consumers are referring to SaaS when they talk about the Cloud, we will begin addressing our concerns at that tier of service. Typically, Cloud users are concerned first about privacy and security. The Cloud is useful for collaborating and sharing data, but what are the risks associated with those benefits? First, there are questions about who controls consumer data. Most consumers do not read the End User License Agreements (EULAs) that they accept in order to use cloud services. Those agreements often indicate that the service provider has more rights to user content than most consumers would willingly accept, if they were aware of the license terms. For example, Facebook's Statement of Rights and Responsibilities states that users grants Facebook

a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use any IP [intellectual property] content that you post on or in connection with Facebook (IP

License). This IP License ends when you delete your IP content or your account unless your content has been shared with others, and they have not deleted it.

(Facebook 2013)

In other words, by posting and sharing content on Facebook, the user is giving Facebook permission to reuse that content. Interestingly, even if the user deletes their account, the data might live on if that content was shared with other users and they have not deleted it; Facebook still claims the right to use that content. While the content is still technically the user's intellectual property, Facebook is free to reuse it.

Users often compromise their privacy and security in ways that are not immediately obvious. One method of gaining unauthorized access to a password-protected account is by social engineering. That is, instead of breaking the password, unauthorized users gather information about users from social networks and other cloud services in order to bypass passwords. For instance, most financial institutions have security challenge questions for recovering a lost password. Typical questions include asking the user's mother's maiden name or the names of relatives or pets. It is possible to glean that information from information the user has herself posted to the Cloud. In 2008, during the U.S. presidential race, vice presidential candidate Sarah Palin's personal email account was subject to unauthorized access when a 20-year old college student used widely available biographical details about Palin to answer challenge questions to reset the password on her account. Additionally, one of the authors of this chapter was once informed by a well-meaning genealogist that he was researching her family tree without her knowledge or input. There was enough information freely available online to trace her family back through several generations.

Questions of unauthorized access can extend into more frightening realms of financial accounts or medical and mental health records. Specifically addressing psychologists while acknowledging the concerns for the medical profession at large, Deveraux and Gottlieb note that:

As records are moved to the cloud, psychologists' ability to exert control over them may diminish to some unknown degree. Such a transition may provide better assurance of electronic record maintenance from disaster such as computer failure or file corruption but could create more difficulties in controlling access. The assurance that no one else can access those records also may be compromised when records are stored in a data center shared by possibly millions of other consumers. Furthermore, the aggregation of sensitive data in such large centers may increase the appeal for potential cybercriminals to steal the information.

(Deveraux and Gottlieb 2012)

In the United States, there are stiff fines levied against health care providers when health records are accessed by unauthorized users; but those fines are small consolation to end users who have their health records appear in Google search results (McCann 2014). Also, while records managers are trained to adhere to data confidentiality policies and laws, not all cloud providers are. Additionally, having a storage provider that is compliant with those data confidentiality laws does not preclude the possibility of a data breach.

We also have concerns about security in the sense that we want to trust that our data in the Cloud will be available when we need it. That is, we have a sense of security about the preservation of our content. However, examples of data loss exist; Google famously lost a few hundred Gmail accounts in 2011 during a software upgrade. While Google was able to restore most of the content of those accounts from backup tapes, losing access to one's data could be catastrophic for small business or consumers.

There is also a false sense of security about the long-term availability of data we store in the Cloud. Once we upload our files, we tend to think that we don't have to do anything else to preserve them. However, the problem of format obsolescence still applies. Previously, when we stored data on disks and hard drives, we had to manage that data every few years to move it onto new media and hard drives as technologies changed. In the process of transferring data from one storage medium to another, we often would catch those times when the software was becoming obsolete, and we would upgrade our files as needed. Now that we put files in the Cloud, we lose those opportunities to check whether we should reformat our files into newer file formats. As any digital archivist can attest, there comes a point where the cost of recovering data stored in obsolete file formats becomes prohibitively expensive, as the hardware and software used to create those files fade into memory.

Another security concern is that, while we store our files in virtual space, those files are, in fact, stored on physical media somewhere. While some have poked fun at consumers who think that weather can impact cloud computing, it can. Power outages from electrical storms have caused data centers to go offline (Miller 2012).

There are other issues with cloud computing beyond privacy and security. The Cloud has become the virtual place where we maintain friendships. There are 'Facebook friends,' which are a class of friends somewhere between friends and acquaintances. Typically, Facebook friends have access to more information about your personal life than acquaintances, but one's allegiance to Facebook friends offline is more akin to that of acquaintances than friends. In some ways, the Cloud has made it easier to maintain long-distance relationships, since we can more easily share photos and communicate with those far away. At the same time, there are concerns that the quality of relationships in the Cloud is inferior to relationships offline.

While an in-depth analysis of friendships in the Cloud is outside the scope of this chapter, there is one important aspect of online relationships we will explore. Many SaaS services, including social networking sites, are free to consumers. The way that these services make money is by mining user data in order to show relevant advertisements to users and by selling consumer data to marketing research companies. Computer scientist and critic Jaron Lanier explains in his book *You Are Not a Gadget* that the entertainment and financial industries have moved away from the idea that "content is king," since it is difficult to reliably predict whether investing in content creation will pay off. Instead, he writes, "the chattering of the cloud itself was a better business bet than paying people to make movies, books, and music" (Lanier 2010: 98). Cloud services that rely on a social component—whether it's the focus of the service, as in Twitter and Facebook, or it's an additional feature of a site with another focus, as in photo-sharing sites such as Instagram and Flickr—are relying on the "chattering of the cloud" for income generation. In a sense, we are selling our social relationships and our communication in exchange for access to these platforms. This is not obvious because we go to these sites to engage in other behaviors, such as sharing photos or posting links to articles and yet, as we engage in these behaviors, we are also selling ourselves and our friendships. Even when we 'like' a product on Facebook, it may not seem immediately obvious to us that we are distributing advertisements to our social network and adding to the market value of that product and of Facebook. It would be strange if we sat down with our friends in a café, pulled out a box of cereal, and told them, apropos of nothing, that we liked this cereal, and yet we do this all the time in social networks.

Additionally, we exchange information about our friends for small gain. It is not merely the content we post and create online that is important for the cloud service; it is the fact that we do this in a network of others that is valuable. Some consumers express concern about the permissions required when installing a free app on a mobile device. As with EULAs, most consumers don't pay close attention to the requested permissions. Apps sometimes request access to the

contacts stored in our mobile devices; the app wants a list of our friends so it can situate us in a larger network of consumers. Imagine if a telemarketer contacted you at home and offered you a few dollars for a copy of your address book. You would probably object to the idea. When we download free mobile apps, we are offered software with a market value of a few dollars, and we willingly turn over our contacts.

Another concern that Lanier raises is that the chattering of the Cloud leads us towards a remix culture where we are not content creators, but rather content sharers and repackagers (Lanier 2010: 121). Content creators are sometimes paid according to clicks on their content, so they are paid for how far their content spreads, not for how good it is, and especially not for how original it is. This apparently has an impact in the offline world, too. How many more ‘relaunches’ of the Batman, Superman, and Spiderman movies do we need? Instead of valuing original content, both producers and consumers seem to get the most out of repackaging content that already exists.

When we move through the tiers of cloud computing, from IaaS, to PaaS, to SaaS, the end users move further and further away from their data and from their computing. We end up with a layer of abstraction where a corporation is between us and our data, and between us and our computing power. Adding the corporation to the mix usually raises privacy concerns, but that is only one problem. GNU founder and free software advocate Richard Stallman criticizes the consumer’s loss of control:

One reason you should not use Web applications to do your computing is that you lose control. It’s just as bad as using a proprietary program. Do your own computing on your own computer with your copy of a freedom-respecting program. If you use a proprietary program or somebody else’s Web server, you’re defenseless. You’re putty in the hands of whoever developed that software.

(Johnson 2008)

This distance between the consumer and the means of production also pushes us from the possibility of being content producers to being content consumers, or, at best, content remixers. Just as most automobile drivers do not understand combustion engines and yet they drive them anyway, SaaS consumers do not understand how computers and the Cloud works, and yet they use them anyway. There is a fundamental difference between driving and using the Cloud: we do not use automobiles to create things. In our heavily digital culture, we *do* use computers to create things. As SaaS users, however, we are moved farther away from those means of production, and that circumscribes our ability to create. We can accomplish only what the software and hardware platforms allow us to accomplish; we can’t change those services to open new horizons of creation. Media theorist and free software proponent Douglas Rushkoff captures this in his command “Program or be programmed.” Rushkoff writes: “Digital technology is programmed. This makes it biased toward those with the capacity to write the code. In a digital age, we must learn how to make the software, or risk becoming the software” (Rushkoff 2010: 134).

A final concern is the Cloud’s tendency to feed into what Michel Foucault describes as a culture of panopticism (Foucault 1995: 195–228). We are immersed in this online culture where we can easily share content and express our approval for content. This in turn shapes our values and behavior, as we are immersed in messages about what other people ‘like’ and what constitutes normal online behavior. (It apparently involves liking lots of cat pictures.) By sharing our content online, and communicating with friends about that content, we open ourselves to constant judgment, and this leads to approval-seeking behavior. One telltale sign of the ubiquity of SaaS in our lives is the way that consumers now go on ‘Facebook fasts’ or ‘social media vacations’ where they deliberately take breaks from being immersed in this milieu.

The Cloud also immerses us in panoptic culture in a more insidious way, in that it provides an incredibly invasive way for governments to spy on citizens, as evidenced in the recent revelations of surveillance conducted by the United Kingdom's Government Communications Headquarters (GHCQ) and the United States' NSA. In George Orwell's *1984*, surveillance was ubiquitous, but comparatively circumscribed when compared to the information available to those who can watch us in the Cloud. Additionally, if the NSA and GHCQ can watch us, some of the same methods that they used can be adopted by cyber criminals to watch us as well.

Cloud Control

Ultimately, issues involving the Cloud take us back to Heidegger's critique of Western metaphysics, culminating as it does in the modern technological age. For Heidegger, modern technology, at its heart, presents the world to us in a particular way. In short, it renders everything it touches as perpetually available, usable, and in many cases disposable, a challenging-forth of beings that reduces the world around us to a collection of resources. At one point, Heidegger illustrates this reduction by showing how a hydroelectric dam fundamentally transforms the Rhine into an energy source (1977: 16), while elsewhere he maintains that nature, at the hands of modern technology, becomes little more than a "gigantic gasoline station" (1966: 50). Western thought, in its effort to fully expose beings to the light of reason, ends up robbing these very beings of their dynamic nature and self-standing.

But what role do we play in this reduction? Are we to blame for this historical unfolding? In a significant sense, not at all, as Heidegger believes that we are challenged as much—or perhaps even more primordially—than the objects with which we engage, pressured as we are to manipulate, commodify, and order the world around us. In this light, the Cloud emerges rather naturally from the modern technological mindset, the supreme danger lying in its tendency to turn us—the alleged users—into a standing reserve in the hands of the Cloud's increasingly few and seemingly insatiable masters.

In contrast to this rather dark, deterministic view of modern technology is a critical theory of technology that emphasizes our role as the makers and shapers of our tools, even in a modern technological age. Andrew Feenberg, for one, thinks that modern technology really does respond to choices made by society, choices that—if made rightly—can go far in realizing and spreading democratic values (Feenberg 1999). The political communities and movements that have formed in recent years through web communication, like those that fueled the Arab Spring and Black Lives Matter, testify to this idea, suggesting that the social networking possibilities afforded by the Cloud can be quite empowering, even amidst some of the real concerns mentioned earlier (Feenberg 2009).

As with any technology, the final verdict on the Cloud—as a tool of liberation or oppression—has much to do with how we comport ourselves to it. To accept what it offers without question or understanding is to reduce ourselves to something of a standing-reserve, as we, *the users*, become *the used*. Our future, however, may be otherwise, so long as we pay genuine attention to the actual workings of the Cloud and its real limitations. Indeed, we must not only be mindful of what the Cloud gives us, but also of the new challenges and moral questions with which it now presents us. Just as Aristophanes's Socrates gazes in wonder at the heavenly Clouds, he may too marvel at our present-day Cloud's transcendent possibilities. Awestruck as he might initially be, however, it would be foolish to think that Socrates would accept the Cloud without question. Lest we allow the Cloud to overwhelm and—dare we say—cloud our powers of moral judgment, questions of how and to what extent it changes our lives for the better must be tempered by questions of how it might change our lives for the worse.

References

- Aristophanes (2000). *Clouds*, trans. P. Meineck. Indianapolis: Hackett.
- Beagrie, N., Charlesworth, A., and Miller, P. (2014). *The National Archives Guidance on Cloud Storage and Digital Preservation: How Cloud Storage Can Address the Needs of Public Archives in the UK*. The National Archives. Available at: www.nationalarchives.gov.uk/documents/archives/cloud-storage-guidance.pdf [Accessed May 30, 2014].
- Devereaux, R. L., & Gottlieb, M. C. (2012). Record Keeping in the Cloud: Ethical Considerations. *Professional Psychology: Research and Practice*, 43(6), 627–632.
- Facebook (2013). Statement of Rights and Responsibilities. Last revision: November 15, 2013. *Facebook*. Available at: www.facebook.com/legal/terms [Accessed May 30, 2014].
- Feenberg, A. (1999). *Questioning Technology*. New York: Routledge.
- Feenberg, A. (2009). Critical Theory of Communication Technology: Introduction to the Special Section. *The Information Society*, 25, 77–83.
- Foucault, M. (1995). *Discipline & Punish*, trans. A. Sheridan. New York: Random House.
- Heidegger, M. (1966). *Discourse on Thinking*, trans. J. M. Anderson and E. H. Freund. New York: Harper & Row.
- Heidegger, M. (1977). *The Question Concerning Technology and Other Essays*, trans. W. Lovitt. New York: Harper & Row.
- Heidegger, M. (2000). *Introduction to Metaphysics*, trans. G. Fried and R. Polt. New Haven, CT: Yale University Press.
- Johnson, B. (2008). Cloud Computing Is a Trap, Warns GNU Founder Richard Stallman. *Guardian.co.uk*, September 29. Available at: www.theguardian.com/technology/2008/sep/29/cloud.computing.richard.stallman [Accessed May 30, 2014].
- Lanier, J. (2010). *You Are Not a Gadget: A Manifesto*. New York: Alfred A. Knopf.
- McCann, E. (2014). Groups Hit With Record \$4.8M HIPAA Fine. *HealthcareITNews*, May 8. Available at: www.healthcareitnews.com/news/group-slapped-record-hipaa-fine [Accessed May 30, 2014].
- Miller, R. (2012). Amazon Data Center Loses Power During Storm. *Data Center Knowledge*, June 30. Available at: www.datacenterknowledge.com/archives/2012/06/30/amazon-data-center-loses-power-during-storm/ [Accessed May 30, 2014].
- Plato (1997). *Phaedo in Plato Complete Works*, ed. J. M. Cooper, trans. G. M. A. Grube. Indianapolis, IN: Hackett.
- Postman, N. (1992). *Technopoly*. New York: Random House.
- Rushkoff, D. (2010). *Program or Be Programmed: Ten Commands for a Digital Age*. Berkeley, CA: Soft Skull Press.

Ethical Issues in Big Data

Deborah G. Johnson

Introduction

In 2014, a paper with the title “Experimental evidence of Massive-Scale Emotional Contagion Through Social Networks” was published in the *Proceedings of the National Academy of Engineering* (Kramer, Guillory, and Hancock 2014). The paper reported on research that had been done using 689,003 Facebook users. Seeking to determine whether the emotional content of users’ postings would be affected by the extent of positive or negative newsfeed the users received, the study involved three groups: one group had their newsfeed filtered for negative content; another group had their newsfeed filtered for positive content; and a third group had their newsfeed randomly filtered. Comparing the postings of the three groups, the researchers concluded that “moods are contagious.” Users who saw more positive newsfeed wrote more positive posts and users who saw more negative content wrote more negative posts. The study contradicted a concern and common belief that when users see a lot of positive content from their friends, it makes them depressed.¹

Publication of the study triggered public controversy. Facebook users were shocked and outraged to learn that they had been used in research without their knowledge or consent. In the debate that ensued, some jumped to Facebook’s defense, citing Facebook’s user agreement. In fact, within the paper, the authors had stated that the research was “consistent with Facebook’s Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research” (Kramer et al. 2014). Others objected to the study on the grounds that it involved manipulation of emotions. A number of internet researchers jumped into the fray with complicated responses expressing their concern that the incident might sour the public on online research and this in turn might lead to tighter restrictions.

In this chapter the controversy surrounding the Facebook emotional contagion study will be used as the starting place for thinking about ethical issues related to big data analytics. Big data analytic techniques are now used to produce knowledge in a broad range of fields of study and in this chapter the focus will be exclusively on knowledge of human characteristics and behavior. After discussing the challenge of framing the emotional contagion study, the chapter addresses issues related to privacy, prediction, and fairness. Big data has enormous potential to generate new knowledge in many different sectors—to contribute to better health, more efficient decision making, better law enforcement, more personalized services, and more. The focus here will be on issues that cut across sectors.

What Is Big Data?

Definitions of the term *big data* vary. They range from those that emphasize the *big*, that is, the enormous scale of data collection and processing, to those that highlight the distinctiveness of

the analytical techniques, to those that identify big data as a cultural phenomenon constituting a new paradigm and way of thinking. A common way of defining big data is to identify its distinguishing characteristics as three Vs: volume, velocity, and variation (McAfee, Brynjolfsson, Davenport, Patil, and Barton 2012). Volume refers to the amount of data, velocity refers to the speed of data analysis, and variety refers to the many types of data that can be combined.

Much of the data about individuals that is now available for analysis has been collected for a purpose other than the one pursued by the big data analysis. For example, in online consumption, personal data may be collected as part of the sales transaction and then used to identify patterns of purchasing behavior among demographic groups for use in future, targeted advertising. Similarly, data about visits to the doctor that are recorded for billing purposes may be used by hospitals and clinics to glean patterns relevant to the efficiency of their operations. From the perspective of data scientists, an enormous amount of data is just sitting around not being used and that data has the potential to provide significant and important information for some purpose. Sometimes new insights are found by combining data from disparate sectors such as looking at the correlation between the distance employees live from their place of work and the employee's longevity in a job or a correlation between those who receive marriage counseling and those who default on loans. An important aspect of big data analysis is that it often involves data moving from one context to another in a way that the individual who provided the data could not have foreseen.

Framing the Emotional Contagion Study

In the public debate about the Facebook emotional contagion study, one of the most striking features of the discourse was the multiplicity of ways in which the study was framed. The study was framed as *scientific research*, *marketing research*, and *emotional manipulation*. At the time of the study, two of the researchers were employed in an academic institution and the paper was published in a prestigious scientific journal, so the study seemed clearly to fall under the category of scientific research. As such, the study would be expected to meet standards for research ethics. Most importantly, since the research made use of human subjects, it would have to undergo scrutiny by an institutional review board (IRB) and the IRB might conclude that the informed consent of the subjects would be required. In fact, the study had been submitted to Cornell University's IRB and it had concluded that no review was necessary. As explained in a statement issued by Cornell after the incident:

Because the research was conducted independently by Facebook and Professor Hancock had access only to results—and not to any individual, identifiable data at any time—Cornell University's Institutional Review Board concluded that he was not directly engaged in human research and that no review by the Cornell Human Research Protection Program was required.

(Carberry, 2014)

Arguably then, when the study is framed as scientific research on human subjects, it does not violate standards for research ethics. The researchers did what they were supposed to do in submitting the study to the IRB, and the IRB concluded that no restrictions were required because the researchers were not working with identifiable personal data. So, what was the problem?

Despite the conclusions of Cornell's IRB, Facebook users were outraged at being used in research without knowing they were participating. In the public debate, some dismissed the outrage by framing the study in another way. Though the term was rarely used, the frame seemed to be that of *internal marketing research*. The claim was that Facebook had the right to do what

it did because users had signed a terms-of-service agreement (often called a user agreement) specifying that Facebook could use information about their users in order to improve its services. The specific phrasing was that Facebook may use information “for internal operations, including troubleshooting, data analysis, testing, research and service improvement” (Hill 2014). Some of the public discussion focused, then, on the adequacy/inadequacy of Facebook’s user agreement and user agreements in general. Although Facebook never conceded that its user agreement was flawed, the company changed the posted agreement not long after the emotional contagion incident (Hill 2014).

The fact that many users were not satisfied with classifying the emotional contagion study as falling under the user agreement is indicative of two important dimensions of the case. First, the study seemed to involve something more than observation. It involved manipulation—and not just any type of manipulation, but manipulation of emotions. This was a third frame under which the study was understood. Manipulation is a complex concept and it might be argued that advertising is a form of manipulation, so some tried to diminish the significance, but the fact that the manipulation in this study involved the moods of users seemed to ignite anger. Although no incidents were reported, the public discussion involved some mention of the possibility that the study might have caused some users to fall into deep depression or experience some other type of harm. This line of argument suggested that Facebook might be entitled to do research that was relevant to improving its service, but that it had stepped over the line when it engaged in research that would have an effect on users’ moods and emotions.

A second important dimension revealed by users’ rejection of the claim that the research fell under the specifications of the user agreement was users’ apparent lack of awareness of how Facebook works. Users’ shock at learning about the emotional contagion study seemed to be due in part, at least, to the surprising realization that Facebook’s algorithms make decisions about what a user sees and doesn’t see. In other words, users seemed to have naturalized Facebook and were jolted by the emotional contagion study into recognizing that the environment in which they operate is intentionally contrived and intentionally contrived by an organization with its own interests. Like other online services such as search engines, the operation of Facebook involves what might be characterized as an upstairs/downstairs setup. Users engage in their lives upstairs—interacting with friends and family online—and their activities are made possible by a massive amount of work that goes on downstairs. Just as the guests at a nineteenth century dinner party might be unaware of the number of cooks and servants that had been involved in the elegant dinner they were being served, Facebook users are generally unaware of the algorithms, software, hardware, and organizational structures that are essential to producing the Facebook user interface. Publication of the emotional contagion study made users aware of the downstairs and the power of decisions made downstairs to affect their experience upstairs.

Thus, the debate about the emotional contagion study illustrates how big data analysis can defy easy categorization. The debate was rich with alternative ways of thinking about an activity that seemed to straddle known categories—scientific research, marketing research, and emotional manipulation. The vying frames in which the study could be put were problematic because ethical principles are often tied to the classification of an activity. Experimentation on humans is treated differently than marketing research and, although both of these may involve manipulation, there are constraints on the kind of manipulation that can be done in either case. The ethical principles that apply to research on human subjects are different from those that apply to marketing research. Research that is observational is treated differently from research that may alter the experiences and behavior of its subjects. Even though the design of a user interface is intended to shape user behavior, design is different from manipulation.

Uncertainty as to how to classify the study was a major factor in the controversy and this uncertainty resulted both from the flow of data across contexts and from the opacity of the data

collection and analysis. The data flowed from a social networking sight to publication in an academic journal. Because the research activity and the normal operations of Facebook were hidden, users were surprised to learn that Facebook contrived and controlled the environment in which they were conducting their social life.

Privacy

The conclusion of Cornell's IRB—that the emotional contagion study was not problematic because the researchers would have no access to personally identifiable information—is a standard way of handling big personal data. The assumption is that when such items as name, address, or social security number are deleted from a data set, privacy is protected. Individuals are, then, thought to be anonymous. Indeed, anonymization is the major strategy by which privacy is thought to be protected in large data analysis projects.

This assumption has, however, been called into question recently because big data analysis tools can be used for reidentification. Arguing that current data collection practices are based on the “deeply flawed” assumption that anonymization protects privacy, Ohm (2010) cites several cases in which enterprising individuals were able to combine publicly released anonymized data with other data to identify individuals. One of the cases involved the then governor of Massachusetts. As the story goes, Group Insurance Commission (GIC)—a Massachusetts government agency that purchased health insurance for state employees—decided to release (to any researcher who requested them) records summarizing every state employee's hospital visits. GIC anonymized the data by removing name, address, and social security number. Ohm explains how Latanya Sweeney was then able to identify the governor's health data:

At the time that GIC released the data, William Weld, then-Governor of Massachusetts, assured the public that GIC had protected patient privacy by deleting identifiers. In response, then-graduate student Sweeney started hunting for the Governor's hospital records in the GIC data. She knew that Governor Weld resided in Cambridge, Massachusetts, a city of fifty-four thousand residents and seven ZIP codes. For twenty dollars, she purchased the complete voter rolls from the city of Cambridge—a database containing, among other things, the name, address, ZIP code, birth date, and sex of every voter. By combining this data with the GIC records, Sweeney found Governor Weld with ease. Only six people in Cambridge shared his birth date; only three were men, and of the three, only he lived in his ZIP code. In a theatrical flourish, Dr. Sweeney sent the governor's health records (including diagnoses and prescriptions) to his office.

(2010: 1719–1720)

So, privacy cannot be dismissed as an ethical issue in big data analyses using anonymized data. Indeed, Ohm argues that there is a tension between the amount of personally identifiable data and the usefulness of the data. The more information there is about individuals, the more useful an analysis is likely to be; the less personal data included, the less useful the analysis. This tension is likely to persist as more and more refined correlations are sought.

Another line of thinking that might also be taken as grounds for dismissing concerns about privacy in big data has to do with a dominant paradigm in understanding privacy. For several decades, privacy scholars used the metaphor of a panopticon to describe and explain the harm of loss of privacy. In the panopticon metaphor, data collection practices mean that individuals are constantly being watched and monitored just as prisoners might be in a panoptic prison.² Although conformity to the rules is a desirable thing in prison, in everyday life being watched all the time has the effect of thwarting freedom and promoting conformity to the norms of

the watchers. Individuals conform to the standards of the watchers for fear of consequences, e.g., being turned down for a job, a loan, insurance, membership in an organization. In this way, surveillance is seen as a threat to civil liberties. For example, one might have great difficulty renting an apartment if one's name appears on a list of tenants who have sued their landlords. Even though citizens have a right to sue landlords who fail to live up to rental contracts, the consequences of pursuing one's right can be costly. In addition to the threat to civil liberties, when people know they are being watched, they may be reluctant to experiment with new, controversial, or deviant ideas, for fear of the consequences. Thus surveillance can have a powerful effect on freedom of expression and the development of new ideas.

The panopticon metaphor has been powerful in understanding concerns about the loss of privacy in modern society. Yet when it comes to big data, the panopticon metaphor does not quite work because individuals are generally unaware of the data collection. When activity is online or through electronic means, data collection is a seamless part of activity. We type or click or press icons with little awareness that records are being created, stored, and analyzed. Since individuals are unaware of being watched, they may not adjust or modify or constrain their behavior. Arguably, then, as long as individuals are unaware of surveillance, the panoptic effect does not occur.

Vaidhyathan (2015) argues that current data collection processes put us in something that is closer to a *cryptopticon*. In a cryptopticon, the conditions in which individuals act are kept hidden (they are, so to speak, encrypted). Still, Vaidhyathan does not conclude that because data collection is hidden, all is well. The cryptopticon is still problematic, and it leads to concerns about the accountability of data collectors. Vaidhyathan (2015) suggests that films and books provide better accounts of how data collection practices are affecting us. They do a better job of taking us towards "the kind of synthetic and ecological understanding we require in order to resist the pernicious effects of massive state surveillance, the dehumanizing aspects of commercial tracking and sorting, and the dangers of social confusion and betrayal."

As an alternative to the panopticon metaphor, Nissenbaum (2009) provides a paradigm for understanding privacy that is well-suited to address the potential harms of big data. Nissenbaum argues that privacy should be understood as contextual integrity. Modern data collection practices violate information norms for particular contexts. Her analysis draws on the fact that in each context in which individuals operate, there are norms with regard to what kinds of information are appropriate/inappropriate and how information will flow. For example, one expects that during a visit to the doctor, the doctor will ask about one's health but not about one's political beliefs or favorite sports teams. One expects that any health information given to the doctor will go into one's medical record and may be used for billing and insurance purposes. However, one would not expect that health information would flow to one's employer or the police or to marketers. Similarly, when one is evaluated at work, one may expect that the evaluation will not be based on one's religious convictions or where one lives. One might expect that the evaluation summary would be shown to higher level administrators, would go into a file, and would be kept for a certain length of time, but one would not expect it to flow to the police or one's spouse or doctor. The point is that privacy has to do with adhering to the information norms for particular contexts. Information norms vary from context to context, and privacy is violated when, contrary to expectations, information appropriate to one context flows to another context in which it is inappropriate.

This flow of information from one context to another is precisely what is made possible and easy with big data analytics. By combining and analyzing databases of information, landlords can learn which tenant-applicants have previously sued their landlord; loan givers can learn whether individuals have sought marriage counseling; employers can learn the financial status and political affiliation of job applicants; health insurers can learn about individual lifestyles.

Personal data that was collected for one purpose is combined with data from another context and used to identify patterns. Decision makers gain new insights and useful information when correlations and patterns are found that help to make better decisions, be it decisions about whom to give a loan, whom to send particular advertisements, whom to hire, whom to insure, and so on. Cross-context correlations constitute predictions of future behavior. To be sure, not all big data analyses involve cross-context correlations, but big data analytic techniques make such analyses possible and easy.

Privacy, Prediction, and Fairness

One of the most publicized cases involving big data predictions and privacy is the case in which a father went into a Target store complaining about the fact that his unmarried eighteen-year-old daughter was being sent coupons for baby clothes and cribs. He accused Target of encouraging her to get pregnant. The daughter received coupons for pregnancy and baby products because she had received a high ‘pregnancy score’ in an analysis of purchasing data. Based on her past purchases, the daughter’s pregnancy score represented the probability that she was pregnant and, therefore, likely to purchase certain kinds of products in the future. Soon after complaining to the manager at the Target store, the father discovered that his daughter was, indeed, pregnant (Hill 2012). The media made much of the fact that Target knew the daughter was pregnant long before the father.

The case is illustrative in part because it challenges the distinction between identifying patterns in categories of people and identifying individuals. Target might have defended itself by saying that it was not interested in the identity of the daughter; it was interested only in nameless groups of individuals who were likely to purchase particular kinds of products in the future. Nevertheless, the information the company collected included names and addresses so that coupons could be mailed. Demonstrating the fragility of privacy, the mere act of sending coupons to a group of individuals resulted, in this case, in revealing highly personal and sensitive information to a specific father about his daughter.

Use of predictive algorithms is now pervasive. Algorithms of the kind that Target used are used to find out if individuals are good credit risks, desirable employees, reliable tenants, likely criminals, and so on. They are used in recommender systems to predict what kind of movies, music, books, and products individuals will purchase. They are used to deliver advertising, to calculate insurance rates, and to determine creditworthiness. They are used to predict how individuals will vote, and the likelihood that an individual will develop particular diseases or do well with a particular kind of surgery. Although predictive algorithms have enormous potential benefit, their use in particular contexts raises a host of ethical issues.

Personalization

‘Personalization’ refers to the activity of delivering advertisements and services that are customized for individuals. In the past the onus might have been on an individual to figure out what they wanted in the way of products, news, or entertainment. Or, individuals might have been asked what their preferences were by means of questionnaires, focus groups, or surveys. Now, however, because so many activities are done online or through electronic media, data about individual behavior are gathered seamlessly and analyzed to determine (predict) what the individual wants *without their consultation*. On the basis of a history of web browsing, online purchases, movie viewing, music downloading, and so on, companies infer individual preferences and interests, and send information accordingly.

Although personalization provides convenience for individuals, ethical concerns have been raised about its more subtle and long-term effects. One such concern is that personalization

leads to each individual living in a *filter bubble*. This is a term coined by Eli Pariser in his 2011 book, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Pariser was struck by the fact that when each of us searches the web using the same term, we each receive different results. Google's algorithms take into account past search histories and in so doing give each individual results that are customized to their interests. Were this just about purchasing products, the concern might not be so great, but the filter bubble also applies to what individuals read and how they get news online. For example, two individuals who read the same online newspaper each day may not be receiving the same collection of stories or at least may not have the stories ordered in the same way. Decisions about what to show an individual are made dynamically on the basis of how the individual has responded to past information, that is, which stories the individual has clicked on in the past.

Concerns about the filter bubble include the worry that individuals will become more and more insular. Instead of being exposed to a diversity of information that might change the way one thinks, one is exposed to information that is likely to reinforce and entrench the views one already has. Personalization seems to treat individuals as already formed and somewhat fixed entities rather than entities that are being shaped by what they are exposed to. The idea of personalization and customization is also somewhat misleading in that more often than not individuals are being shaped to fit into the categories of the personalization system. For example, when—based on one's profile—one is offered a filtered set of purchasing options or a set of news articles to read, the convenient thing to do is to choose among the given options. The record of what one chooses then goes back into one's personal profile reinforcing that one is a person who fits the category even though if the individual had been given a wider range of options, the individual might have chosen something else. In this way, personalization intensifies one's fit with a category and, in a sense, makes the person fit the category.

Because personalization narrows the range of information to which individuals are exposed, some have argued that personalization leads to fragmentation, especially social and political fragmentation (Coudry and Turow 2014). For example, when progressives and conservatives are exposed only to news that fits their ideology, they will be less familiar with and may become less tolerant of other ideologies. So the insularity of personalization also means that individuals in particular groups may become more and more similar while at the same time becoming more and more different from other groups. Since democracy depends at some level on shared beliefs, intensification of fragmentation is a threat to democracy.

Whether or not personalization causes fragmentation is to some extent an empirical question. A number of scholars have begun to study this topic, and a body of research is building but is not yet definitive. It seems to present a mixed picture of how personalization can be used to lead to more commonality (Hosanagar, Fleder, Lee, and Buja 2013). Some of the research considers alternative ways to design personalization systems so as to expose individuals to diverse perspectives and avoid the pitfalls of fragmentation (Garrett and Resnick 2011).

Whether or not personalization leads to fragmentation, the opacity of personalization systems seems problematic. The controversy over the emotional contagion study suggests that users tend to naturalize online environments, failing to recognize the extent to which filtering is happening and unaware of the way in which the interests of the system owner may shape the environment. Thus, more transparency has become a common call of those who study personalization. The idea is that more transparency would increase consumers' awareness of the ways in which information is being filtered by news providers, search engines, and advertisers.

Fairness

One of the most disturbing aspects of big data analytics is the potential for unfair practices and indirect discrimination. As data flows freely from one sector to another, organizations are able

to make decisions about individuals on the basis of information that in the past was unavailable and thought to be irrelevant. Instead of basing decisions on preconceived notions of what is relevant, decision makers can let data analysis show them what is relevant. For example, if credit card data combined with loan default information shows a pattern of higher probability of loan default among those who use go to marriage counseling, then companies that provide loans may change their decision-making models to take into account whether an applicant has gone to marriage counseling recently. Such a practice has implications that are far broader than lenders' interests in managing their risk.

For one, making it harder for couples to get loans might exacerbate their marital problems, leading to a split-up that would create or worsen financial problems and thereby contribute to the self-fulfilling prophecy of loan default.³ Even if this were not so, however, individuals who apply for loans do not expect the fact that they have gone to marriage counseling to affect their credit rating. This seems a clear case of contextual integrity being violated. In this case, the violation need not be due to a marriage counselor revealing information about clients, but due to use of a credit card to pay for the counseling visits. Although one might expect the balance on one's credit card to impact one's credit rating, one doesn't expect the particular purchases one makes to be used in other contexts. Among other problems with the use of marriage counseling to determine loan eligibility, if such a practice were known to the public, fewer couples might seek marriage counseling fearing the consequences for their credit.

The Upturn Report (Robinson, Yu, & Reinke 2014) identifies a wide range of concerns about how big data are used in decision making in the financial sector, employment, government, and criminal justice. One of the examples the Report uses is a practice in which insurance companies use big data to infer one's health condition. The Report describes employees at Deloitte (a highly prominent accounting firm) explaining that they can "predict a life insurance applicant's health status with an accuracy comparable to a medical exam" (Robinson et al. 2014, n.p.). When institutions use data to indirectly infer one's health condition, the potential for discriminatory practices increases enormously. Even though insurance companies may be prohibited from denying insurance to individuals on the basis of preexisting conditions, now they can pick and choose among customers based on probabilities of acquiring diseases in the future, or they can charge more to those with higher probabilities of acquiring certain diseases. Decision making based on probabilities has an inherent problem in that it means that some will pay more and never get the predicted medical condition.

Putting aside the inherent problem with prediction, when big data analysis shows that some factor is relevant, the correlation may be the indirect result of racial or gender categories. An interesting illustration of this is found in insurance, where models show that those who live far from work and drive at night are more likely to have automobile accidents. The night driving correlation may be due to partying, late attendance at bars, and alcohol consumption; however, since low-income individuals and especially people of color are more likely to work nightshifts and to live farther from work, their insurance rates may be higher (Robinson et al. 2014: 6). Thus, what might seem like a neutral correlation will, if used, reinforce discrimination.

Employers have discovered a correlation between the distance an employee lives from work and staying at a job; that is, individuals with shorter commutes tend to be better hires because they stay in their jobs longer. The Upturn Report suggests that this correlation may have a negative impact on those who can't afford to live close to work:

Such a preference punishes people for living far from where the jobs are, and can particularly hurt those living in economically disadvantaged areas, who are disproportionately people of color. Such practices make it even harder for people in disadvantaged communities to work their way out of poverty.

(Robinson et al. 2014: 15)

This example has an effect like that of personalization in that decisions based on a correlation can reinforce and entrench a pattern that already exists.

Similar concerns arise with the use of big data in the context of criminal justice. Use of big data by police is referred to as ‘predictive policing.’ Police in Chicago have used data analysis to identify those who are most likely to be involved in ‘future violence.’ In other places, fore-closure data are used to predict where crimes are most likely to occur, and key word analysis of Twitter feed is used to develop analytics for predicting crime (Joh 2014). One writer summarized the potential harm as follows:

Critics say that surveillance technologies carry risks to civil liberties and that predictive policing, in particular, can additionally perpetuate a self-fulfilling cycle of bias. That is to say, an area with historically high rates of crime gets greater police attention, which results in more arrests, which in turn the algorithm uses to deem that neighborhood an area where crime is more likely to occur.

(Sengupta 2013)

Depending on how big data are used in policing, they have the potential to address crime in a preventative way—but at the same time this poses ethical challenges that go to the heart of a fair criminal justice system.

Conclusion

Collection and use of personal information is not new, but big data analytics make it possible for data to move without constraints from one context to another and for information about individuals to be used in decision making regardless of how the information was acquired and regardless of the implications. Currently, there is a free-for-all in the sense that there are few constraints on what information can be used and little attention being given to the changes being made in contextual information norms. The ethical challenges are nothing short of daunting because big data threaten to change fundamental relationships between public and private organizations *and* individuals (consumers and citizens) by changing information norms. Although the new information norms do not constitute a panopticon, they give enormous power to data collectors and decision makers to serve their own interests and not necessarily the interests of consumers and citizens or the interests of a democratic society. Big data practices are not transparent and there are few mechanisms for holding institutions accountable for their use of big data. Requiring more transparency and more accountability would go a long way towards addressing the ethical issues and ultimately ensuring that the benefits of big data analytics can be had without undermining privacy and fairness.

Notes

- 1 James Grimmelman has documented the case in a website that contains the original paper together with everything that was written about it up to June 30, 2014. See: http://laboratorium.net/archive/2014/06/30/the_facebook_emotional_manipulation_study_source [Accessed May 2, 2016]
- 2 The idea of the panopticon originates with Jeremy Bentham’s design for a prison in which the prison cells were to be arranged in a circle around a guard tower. The inside wall of the cells faced the guard tower so that prisoners could be observed at all times. Michel Foucault later developed ideas about how the prison design works to explain broader phenomenon in modern society.
- 3 According to *Federal Trade Commission v. CompuCredit Corporation and Jefferson Capital Systems* (2008), CompuCredit is said to have used transactions for marriage counseling and personal counseling as well as transactions with massage parlors, bars and nightclubs, and pool and billard halls in its

calculation of credit scores. See www.ftc.gov/enforcement/cases-proceedings/062-3212/compucredit-corporation-jefferson-capital-systems-llc [Accessed May 1, 2016].

References

- Carberry, J. (2014). Media Statement on Cornell University's Role in Facebook 'Emotional Contagion' Research. Media Relations Office, Cornell University. Available at: <http://mediarelations.cornell.edu/2014/06/30/media-statement-on-cornell-universitys-role-in-facebook-emotional-contagion-research/>
- Couldry, N., and Turow, J. (2014). Advertising, Big Data and the Clearance of the Public Realm: Marketers' New Approaches to the Content Subsidy. *International Journal of Communication*, 8, 1710–1726.
- Garrett, R. K., and Resnick, P. (2011). Resisting Political Fragmentation on the Internet. *Daedalus*, 140(4), 108–120.
- Hill, K. (2012). How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did. *Forbes*, February 16. Available at: www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#e91696e34c62 [Accessed April 18, 2016].
- Hill, K. (2014). Facebook Added 'Research' to User Agreement 4 Months After Emotion Manipulation Study. *Forbes*, June 30, 2014. Available at: www.forbes.com/sites/kashmirhill/2014/06/30/facebook-only-got-permission-to-do-research-on-users-after-emotion-manipulation-study/#26640bb910c1 [Accessed April 30, 2016].
- Hosanagar, K., Fleder, D., Lee, D., and Buja, A. (2013). Will the Global Village Fracture Into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation. *Management Science*, 60(4), 805–823.
- Joh, E. E. (2014). Policing by Numbers: Big Data and the Fourth Amendment. *Washington Law Review*, 89, 35–68.
- Kramer, A. D., Guillory, J. E., and Hancock, J. T. (2014). Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., and Barton, D. (2012). Big Data: The Management Revolution. *Harvard Business Review*, 90(10), 61–67.
- Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press.
- Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57, 1701.
- Pariser, E. (2011). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin.
- Robinson, D., Yu, H., and Rieke, A. (2014). Civil Rights, Big Data, and Our Algorithmic Future. *Robinson+Yu*.
- Sengupta, S. (2013). In Hot Pursuit of Numbers to Ward Off Crime. *Bits*, June 19, 2013. Available at: http://bits.blogs.nytimes.com/2013/06/19/in-hot-pursuit-of-numbers-to-ward-off-crime/?_r=1 [Accessed May 1, 2016].
- Vaidhyanathan, S. (2015). The Rise of the Cryptopticon. *Hedgehog Review*, 17(1), 72–85.

Virtual Environments

Johnny Hartz Søraker

Philosophy has been preoccupied with notions like ‘truth,’ ‘reality,’ and ‘reference’ since the beginning, so it should come as no surprise that the notion of ‘virtual’ entities and environments raises a host of epistemological, ontological, and ethical challenges. The term *virtual* is used in a myriad of ways, including the generic notion of being ‘almost but not quite’ (e.g., virtually impossible), the engineering sense of simulating a piece of soft- or hardware (virtual desktops, virtual memory), and the philosophical sense of being ‘quasi-,’ ‘pseudo-’ or ‘almost the same as’; something that is almost but not quite real, or something real without being *actual* (cf. Shields 2003: 25, inspired by Proust and Deleuze). In the same vein, Jaron Lanier, one of the early pioneers in virtual reality (VR) research who also popularized the term *virtual reality* explains that for something to be virtual “it has to be indistinguishable [from the actual entity] in some practical context, while it remains distinguishable in another” (Lanier 1999).

This chapter will focus on the philosophical issues raised by computer-generated, interactive environments that give users the illusion of perceiving a three-dimensional world consisting of virtual objects that can be interacted with by the means of computer peripherals. Although there is no consensus, *virtual environment* is typically used in this broad sense, and subsumes *virtual worlds*, which refers to persistent virtual environments in which users are represented as avatars and can interact with each other. *VR* is a closely related and perhaps better-known term, but is typically reserved for virtual environments in which the user’s field of vision is substantially replaced by computer-generated visual output and in which interactions with the virtual environment, including the user’s viewing angle, are determined by tracking the movement of several body parts in real time. Hence, VR requires either a head-mounted display or a surrounding projection screen, as well as peripherals that detect real-time bodily motions, such as a data suit, glove, and/or helmet. Virtual environments, especially in the form of video games, are more commonly realized on regular computer screens and interacted with by means of standard computer peripherals such as a mouse, keyboard, or game controller. The pace of technological development in these fields makes it difficult to maintain clear distinctions, however, especially with the advent of VR peripherals such as *Oculus Rift* currently being targeted at the consumer market.

According to Michael Heim’s early analysis of the metaphysics of VR, there are seven inter-related concepts that are central to any kind of virtual technology: simulation, interaction, artificiality, immersion, telepresence, full-body immersion, and networked communication (Heim 1993: 109–115). It is important to note that none of these are, on their own, necessary or sufficient conditions for something to be virtual. Virtual environments can be used to simulate real environments, such as existing buildings or city areas, or to visualize imaginary ones, for instance spaceships or battlegrounds. VR is capable of delivering the most realistic experience, hence it is often used for realistic immersion that would be practically, economically, or ethically difficult to undertake in the real world, such as skills training (e.g., military or medical

operations) computer-aided design (e.g., three-dimensional blueprints), and investigation of objects that are inaccessible to the human eye in the real world (e.g., molecules, internal organs, and galaxies). Pending further advances in network speed and bandwidth, the massive amounts of data and computational resources needed for genuine VR entails that it rarely involves multiple users interacting in real-time at a distance. Although the terms cover a large range of different uses and features, any type of virtual environment—which subsumes VR and virtual worlds—must be computer-generated and interactive. This excludes computer-generated environments that are not interactive (e.g., three-dimensional cinema), as well as interactive technologies that do not have computer-generated visual output.

Ontological Issues

Virtual environments and real environments can be seen as opposite ends of a continuum, with augmented reality in the middle—a form of virtual environment in which the user's view of the physical world is supplemented by virtual entities (Milgram et al. 1995). This idea is also captured by Baudrillard's notion of "hyperreality" as "the generation by models of a real without origin or reality" (Baudrillard 1994: 1), echoing the common interpretation of the virtual as denoting a new, fictional kind of reality—as something "almost but not quite real." This is unfortunate and can give rise to the mistaken belief that virtual actions do not have real consequences, dubbed the "virtuality fallacy" (Tavani 2010: 97). This can, in turn, lead to an impoverished understanding of the effects that one's words and actions may have on other users when mediated by a virtual world.

The difference between real and virtual is inherently difficult to clarify, due to four interrelated reasons. First, virtual environments are made possible by a physical computer that runs the simulation, which entails that even if the virtual entities and environments as such are not physical, they do exist in the form of a digital representation in a physical medium. Second, reality is partly socially constructed (Searle 1995), and many social constructions can be ontologically reproduced in a virtual environment without loss of properties (Brey 2003). For instance, a piece of paper with a particular set of characteristics may count as a dollar bill in the context of a particular jurisdiction, in just the same way as currency in a virtual environment. Third, from a Platonic point of view one may consider virtual environments as being mere depictions and hence further removed from ultimate reality, but virtual entities are often constructed out of a mesh of triangles and other mathematically defined geometric figures, hence could be argued to be *more* ideal than physical entities. Finally, a virtual environment is sometimes referred to as being surrounded by a 'magic circle' that delimits the actions and experiences that form part of the virtual environment and those that do not. However, as Edward Castronova (2005) points out, this is probably better described as a "membrane," which allows for the transfer of behavior, attitudes, beliefs, and desires from one to the other. Since virtual environments therefore may be as real as the real world, Albert Borgmann (1999) suggests a distinction between *virtual* and *actual* instead, the latter referring to the physical universe inhabited by our biological bodies. This account presupposes that actual reality is the primary reality, the reality in which meaning must come from, but this is only natural, according to Borgmann, because the "human body with all its heaviness and frailty marks the origin of the coordinate space we inhabit . . . this original point of our existence is unsurpassable" (Borgmann 1999: 190). Robert Nozick indirectly makes a similar point in his famous thought experiment, "the experience machine." Nozick argues that if there were a VR in which *all* our desires would be satisfied (disregarding any practical concerns), most of us would still not plug in and leave our physical lives behind. The reason, Nozick claims, is that such a simulation would have no "actual contact with any deeper reality" (Nozick 1993: 43). In this thought experiment, Nozick captures a common

concern when it comes to our actions in virtual worlds: since virtual actions are not directed at actual entities, they are inauthentic and *ipso facto* less valuable.

We may expect the blurring of virtual and actual to become even more complex with the advent of virtual environments that are blended with actual reality (augmented reality, e.g., Google Glass and Microsoft Hololens), as well the generations of people coming of age who grew up with the two worlds constantly overlapping each other—so-called digital natives (Prensky 2001). The question of whether or not we can or should distinguish between the two is important, however, since many of the ethical issues surrounding virtual environments revolve around the question of their reality.

Epistemological Issues

To illustrate the complexity of epistemology in relation to virtual worlds, it is helpful to consider how virtual environments (fail to) correspond to standard accounts of technological mediation. In his influential *Technology and the Lifeworld* (Ihde 1990), Don Ihde conceptualizes four different ways in which technologies mediate between humans and the world. First, some technologies become *embodied*, meaning that the technology in question alters the way in which we perceive the world without the technology itself being explicitly present: the technology “disappears” into the background when it is being used. Typical examples include glasses, microscopes, and telescopes, which allow us to perceive parts of the world we would not otherwise see, without noticing the technologies that make this possible. Second, technologies can form an *alterity* relation, in which we interact with the technology itself as an Other—leaving *the world* more or less in the background. Typical examples of alterity relations include the withdrawal of money from an ATM or interacting with a robot. Third, some technologies form part of a *hermeneutic* relation, in which a part or feature of the world can be read (and in some cases interpreted) by human beings by means of a technology. A standard example given for such a relation is the thermometer, which “hermeneutically delivers” a representation of a particular aspect of the world (Ihde 1990). Finally, Ihde also introduces *background* relations, wherein the technology is not perceived directly but becomes “a kind of near-technological environment itself” (Ihde 1990: 108). The most obvious examples include the kinds of technologies that surround us in everyday life, such as technologies for lighting, heating, air conditioning, etc.

When we try to apply these relations to virtual worlds and entities, we quickly find ourselves in trouble—but these problems are in themselves interesting because they reveal part of what is so unique about virtual worlds. One initial problem is that the distinction between “technology” and “world” featured in Ihde’s relations becomes complicated when we talk about virtual worlds. We might propose to have “world” refer exclusively to the *physical* world, but this leads to a couple of peculiar problems. First, since we use technologies (computers and peripherals) to access virtual worlds, are these technologies part of the mediation between the human being and the virtual world? Second, how do we conceptualize technologies *within* the virtual world itself? If we look at the examples given earlier, it seems like all of Ihde’s relations can be applied to virtual entities *within* virtual worlds as well: there are virtual glasses, virtual meters of various sorts, virtual ATMs, and virtual lighting. Are these entities part of the virtual-world-as-technology, or are they themselves technologies within another technology? It could be argued that we do not directly interact with these technologies, but reducing a virtual world to our interaction with the computer and peripherals hardly makes sense.

To make matters even worse, the virtual world can be seen as a form of hermeneutic technology as well, because it does mediate the physical world—in particular the actions of *other* human beings. To illustrate this, take a regular computer game that does not allow for online multiplayer interactions. Such games can be considered as a form of *alterity* relation, because

we interact with the technology while the world is more or less in brackets. With virtual worlds (i.e., when there is an online multiplayer element), the world is no longer bracketed because we communicate with actual human beings through the virtual world/technology. This, in turn, is seamlessly combined with all kinds of *nonhuman* ‘alters’ within the virtual world. Finally, as if the notions of ‘technology’ and ‘world’ are not difficult enough to place in these relations, the concept of ‘human’ is also complicated by the fact that the interactions are carried out *as if* done by a representation of the human (an *avatar*), and done from the standpoint of that avatar’s indexical location within the world/technology.

In short, with virtual worlds the relations in the technology relations become ambiguous. ‘Human’ may refer to the actual human or the avatar representation. ‘Technology’ may refer to the user’s computer and peripherals, the computer simulation and databases that underpin the virtual worlds, the virtual world itself (as experienced), and/or virtual entities within virtual worlds. Finally, ‘world’ can refer to the actual or the virtual world. These problems show that virtual worlds and entities probably cannot be conceptualized in the same manner as many, if not most, other technologies. Virtual worlds are both worlds and technologies; the computer simulation is both the underpinning of the virtual world and the means of mediation; entities within virtual worlds can be regarded as technologies themselves—and although virtual worlds mediate the physical world and other human beings, they *also* construct reality. All of this complexity shows how epistemologically unique virtual worlds are, and how difficult it is to conceptualize the relations between humans, virtual worlds, and the physical world.

Ethical Issues

These epistemological and ontological issues have direct bearing on the ethical issues, since many of the most common points of criticism rest on a presumed difference between virtual and actual, and a corresponding inferiority in value. According to this line of reasoning, various virtual ‘surrogates’ lack something found in the actual counterpart, and this lack translates to reduced value—and because of this we *ought* to choose the real over the virtual. This is a particularly common strategy when it comes to assessing the value of virtual relationships and communities. Hubert Dreyfus is one of many philosophers who have criticized virtual worlds as inferior since we cannot have the same range of movements and expressions of the body in virtual worlds, nor a sense of context, commitment, or shared risk-taking. Furthermore, the lack of physical context fosters what he characterizes as the nihilist, irresponsible, and often uninformed nature of virtual communities (Dreyfus 2004). Albert Borgmann, on the basis of a distinction between instrumental, commodified, and final communities, argues that virtual communities and relationships can at best be instrumental or commodified, because they do not contain “the fullness of reality, the bodily presence of persons and the commanding presence of things” found in final communities (Borgmann 2004: 63). In a similar fashion, Darin Barney (2004) sees virtual communities as inferior due to their lack of physical *practices*, and Howard Rheingold argues that the lack of spontaneous bodily gestures and facial expression is the reason for the “ontological untrustworthiness” of virtual acts of communication (Rheingold 2000: 177). A related argument has also been made by Langdon Winner, who argues that virtual communities ought not to be regarded as communities at all, because this ignores the importance of “obligations, responsibilities, constraints, and mounds of sheer work that real communities involve” (Winner 1997: 17).

More directly to the value of friendship in virtual worlds, Cocking and Matthews (2000) argue that virtual friendships currently do not allow for nonvoluntary self-disclosure, and that genuine friendships can be established only on the basis of nonvoluntary self-disclosure. According to this account, the only way to genuinely know someone is to spend considerable

amounts of time in their physical presence. This approach to investigating the value of virtual relationships in particular, and computer-mediated communication in general, is often referred to as the “cues filtered-out” approach (cf. Joinson 2003: 25–37).

In a closely related argument, Shannon Vallor holds that virtues like patience, honesty, and empathy may be threatened in virtual worlds. The key problem in this regard is that virtual environments are often designed with the aim of giving users the possibility to engage in otherwise unavailable or difficult human activities, which often run counter to “the fact that virtues are typically developed as a consequence of performing actions that are, at least initially, difficult” (Vallor 2010: 168). Another problem with the lack of embodied Others in virtual worlds is that we are not reminded of the radical personal and cultural differences between us: “The absence of the embodied self and its multiple cues of radical difference in online contexts make it easier for us to behave in ethnocentric and thereby imperialistic fashion” (Ess 2009: 115).

The extent to which we can differentiate between real and actual also forms the basis of several ethical issues. Some authors, such as Michael Heim (1993) and Sherry Turkle (1995), have argued that a distinction between physical and virtual will always exist because people are biological human beings that are born and die in the physical world and retain their roots there. Philip Zhai (1998), on the contrary, has argued that such biological background facts are irrelevant and that future, multiuser VR can offer us a limitless world as rich and detailed as physical reality. Zhai further argues that VR can be augmented with peripherals that can take care of any biological function, even sexual reproduction, thus being able to replace the physical world as one’s primary habitat.

There are also several negative social consequences that could result from extensive use of virtual environments, especially insofar as the supposedly idealized, vacuous, and consequence-less virtual environments come to serve as a model by which people comprehend the actual world, and conversely, that the attention and care that we attach to real-world people, animals, and things is also attached, inappropriately, to virtual entities and characters. Another worry is that people may come to prefer the freedom and limitlessness of virtual environments over the limitations of physical existence and invest most of their time and energy in their virtual life, to the neglect of the real people and affairs in their physical lives. Indeed, there is much controversy surrounding the question of using virtual environment as a way of escaping one’s real-life problems as well as the question of virtual environments and addiction. ‘Internet Gaming Disorder’ is currently identified as a condition warranting more clinical research before it can be included in the Diagnostic and Statistical Manual of Mental Disorders (Petry et al. 2014).

It is sometimes claimed that since virtual environments are not real, the consequences of one’s actions in virtual environment are not real-life consequences. However, since the computer simulation that underpins the virtual environment is a physical entity capable of triggering real-life consequences, it is better to distinguish between “intravirtual” and “extravirtual” consequences (Søraker 2012). Every action we perform in a virtual environment can be construed as consisting of both intravirtual and extravirtual consequences. Intravirtual consequences affect only the state of the virtual environment, whereas extravirtual consequences are triggered by the state of the virtual environment yet have potentially dramatic consequences in the real world. The latter includes all kinds of physical events that can be triggered by the computer simulation, including the way in which the computer-generated output affects users’ beliefs, desires, and emotions. It is sometimes possible to perform actions in virtual environment that would be cruel and immoral in the real world, but can be performed without retribution in virtual environment because there is supposedly no real harm done. This is often true, but might lead us to forget that altered behavioral dispositions and emotional reaction to offense is also a real consequence. Furthermore, we may ask whether it is morally defensible for people to act out graphic and detailed scenarios of mass murder, torture, and rape in virtual environment, even when done in

private. Are there forms of behavior that should not be morally or even legally acceptable even in virtual environment, either because of their intrinsically offensive nature, or because such simulations desensitize individuals and may facilitate immoral behavior in the real world? Or is it the case that the possibility to act out fantasies in virtual environment keeps some people, such as sex offenders or people prone to violence, from acting out this behavior in the real world, so that virtual environment may actually prevent crime? The answers to such questions are crucially determined by which ethical perspective we adopt. From a virtue ethics point of view, violent and offensive behavior in virtual worlds may be regarded as unethical insofar as it “erodes one’s character and makes it more difficult for one to live a fulfilled eudaimonic life” (McCormick 2001: 277). From a utilitarian point of view, however, such behavior is deemed unethical only if a causal link between virtual actions and real actions can be empirically demonstrated. Deontological ethics is notoriously difficult to apply to single-user virtual environment, but notions like respecting people’s humanity becomes relevant for how we treat others, via their avatars, in virtual worlds.

The extent to which virtual environments represent actual reality is not only relevant for epistemological and ontological questions, but also gives rise to several ethical problems. virtual environments that are intended to simulate actual realities may misrepresent these realities, according to expected standards of accuracy, especially with VR where the expectation of realism is much higher than for virtual environments in general. This may cause users to make false decisions or act wrongly, with potentially serious consequences, especially in areas in which life-or-death decisions are made, such as medicine and military combat. When VR is used for education and training, therefore, developers have a responsibility to adhere to high standards of accuracy and realism (Brey 1999). Virtual environments—video games in particular—may also contain biased representations that are not necessarily false, but that contain prejudices about people or situations. For example, a virtual environment may represent minorities in stereotypical ways, or a combat simulation may simulate combat situations only in which civilians are absent. Like other media, virtual environments may also break taboos by depicting morally objectionable situations, including violent, blasphemous, defamatory, and pornographic situations. This has particularly been an issue with virtual worlds, in which users can simulate acts often regarded as offensive, whether willfully or forced. To name just a few examples, there has been much controversy surrounding phenomena like virtual pedophilia and virtual rape (Dibbell 2007). These phenomena are not only challenging to deal with for ethicists, but also challenge many traditional definitions and categories in (philosophy of) law (Strikwerda 2013).

Other ethical issues relate to identity, particularly when it comes to interaction between multiple users in virtual worlds. Role-playing in virtual worlds enables people to experiment with identities and to experience Otherness more vividly than ever before. Although the parallel is highly questionable and ignores the complexity of real-life stereotypes, someone may portray themselves as the opposite gender or a different ethnicity, potentially giving some understanding of what it is like to be the other (Nielsen 2015). So-called VR exposure therapy has also been used successfully to treat anxiety and specific phobias, such as overcoming agoraphobia by interacting with others in a virtual world, or arachnophobia by interacting with virtual spiders (Parsons and Rizzo 2008). Sherry Turkle (1995) made an early argument to the effect that the freedom to portray multiple identities decoupled from our bodily appearance helps us

develop models of psychological well-being that are in a meaningful sense postmodern. . . . They acknowledge the constructed nature of reality, self, and other. . . . We are encouraged to think of ourselves as fluid, emergent, decentralized, multiplicitous, flexible, and ever in process.

(Turkle 1995: 263–264)

In later writings, however, Turkle (2012, 2015) has become much more skeptical of this, advocating instead increased importance attached to real social contact and undivided attention.

In conclusion, there is little doubt that virtual environments will continue to pose challenges in all areas of philosophy—and that these challenges will change in line with technological developments that allow for more complex and lifelike behavior and experiences. With the number of users rising steadily and demographics constantly changing, especially when it comes to gamers' gender and age (Entertainment Software Association 2015), it is also likely that these questions will only become more pressing and ubiquitous in the near future. As Edward Castronova points out, “should the number of people spending most of their time in the [virtual] world become quite large—and all I am willing to say at this point is that it is plausible, not certain—their decisions will impose a paradigm shift on everyone” (Castronova 2005: 277).

References

- Barney, D. (2004). The Vanishing Table, or Community in a World That Is No World. In A. Feenberg and D. Barney (eds.) *Community in the Digital Age: Philosophy and Practice*. Oxford: Rowman and Littlefield, 31–52.
- Baudrillard, J. (1994). *Simulacra and Simulation*. Ann Arbor, MI: University of Michigan Press.
- Borgmann, A. (1999). *Holding On to Reality: The Nature of Information at the Turn of the Millennium*. Chicago: University Of Chicago Press.
- Borgmann, A. (2004). Is the Internet the Solution to the Problem of Community? In A. Feenberg and D. Barney (eds.) *Community in the Digital Age: Philosophy and Practice..* Oxford: Rowman and Littlefield, 53–68.
- Brey, P. (1999). The Ethics of Representation and Action in Virtual Reality. *Ethics and Information Technology*, 1(1), 5–14.
- Brey, P. (2003). The Social Ontology of Virtual Environments. *American Journal of Economics and Sociology*, 62(1), 269–282.
- Castronova, E. (2005). *Synthetic Worlds: The Business and Culture of Online Games*. Chicago University of Chicago Press.
- Cocking, D., and Matthews, S. (2000). Unreal Friends. *Ethics and Information Technology*, 2(4), 223–231.
- Dibbell, J. (2007). *My Tine Life: Crime and Passion in a Virtual World*. Available at: www.lulu.com/content/1070691 [Accessed January 1, 2016].
- Dreyfus, H. L. (2004). Nihilism on the Information Highway: Anonymity Versus Commitment in the Present Age. In A. Feenberg and D. Barney (eds.) *Community in the Digital Age: Philosophy and Practice*. Oxford: Rowman and Littlefield, 69–82.
- Entertainment Software Association. (2015.) *Essential Facts About the Computer and Video Game Industry*. Available at: www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf [Accessed September 1, 2016].
- Ess, C. (2009). *Digital Media Ethics*. Cambridge: Polity Press.
- Heim, M. (1993). *The Metaphysics of Virtual Reality*. Oxford: Oxford University Press.
- Ihde, D. (1990). *Technology and the Lifeworld*. Bloomington, IN: Indiana University Press.
- Joinson, A. N. (2003). *Understanding the Psychology of Internet Behaviour: Virtual Worlds, Real Lives*. New York: Palgrave-MacMillan.
- Lanier, J. 1999. *Virtual Reality*. Available at: http://findarticles.com/p/articles/mi_m0GER/is_1999_Fall/ai_56457593 [Accessed January 6, 2016].
- McCormick, M. (2001). Is It Wrong to Play Violent Video Games? *Ethics and Information Technology*, 3(4), 277–287.
- Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum. In Hari Das (ed.) *Proceedings of SPIE–The International Society for Optical Engineering*. Boston, MA: SPIE, 282–292.
- Nielsen, D. (2015). Identity Performance in Roleplaying Games. *Computers and Composition*, 38 (Part A), 45–56.

- Nozick, R. (1993). *Anarchy, State and Utopia*. Oxford: Blackwell.
- Parsons, T. D. and Rizzo, A. A. (2008). Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-Analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3), 250–261.
- Petry, N. M., Rehbein, F., Gentile, D. A., Lemmens, J. S., Rumpf, H.-J., Mößle, T., Bischof, G., Tao, R., Fung, D. S. S., Borges, G., Auriacombe, M., González Ibáñez, A., Tam, P., and O'Brien, C. P. (2014). An International Consensus for Assessing Internet Gaming Disorder Using the New DSM-5 Approach. *Addiction*, 109(9), 1399–1406.
- Prensky, M. (2001). Digital Natives, Digital Immigrants Part 1. *On the Horizon*, 9(5), 1–6.
- Rheingold, H. (2000). *The Virtual Community: Homesteading on the Electronic Frontier*. Cambridge, MA: MIT Press.
- Searle, J. (1995). *The Construction of Social Reality*. New York: Free Press.
- Shields, R. (2003). *The Virtual*. London: Routledge.
- Søraker, J. H. (2012). Virtual Worlds and Their Challenge to Philosophy: Understanding the 'Intravirtual' and the 'Extravirtual'. *Metaphilosophy*, 43(4), 499–512.
- Strikwerda, L. (2013). When Should Virtual Cybercrime Be Brought Under the Scope of the Criminal Law? In M. Rogers and K. Seigfried-Spellar (eds.) *Digital Forensics and Cyber Crime*, vol. 114. Berlin Heidelberg: Springer, 109–143.
- Tavani, H. T. (2010). *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing*. Danvers, MA: John Wiley & Sons.
- Turkle, S. (1995). *Life on the Screen: Identity in the Age of the Internet*. New York: Simon & Schuster.
- Turkle, S. (2012). *Alone Together: Why We Expect More From Technology and Less From Each Other*. New York: Basic Books.
- Turkle, S. (2015). *Reclaiming Conversation: The Power of Talk in a Digital Age*. New York: Penguin.
- Vallor, S. (2010). Social Networking Technology and the Virtues. *Ethics and Information Technology*, 12(2), 157–170.
- Winner, L. (1997). Cyberlibertarian Myths and the Prospects for Community. *SIGCAS Computers and Society*, 27(3), 14–19.
- Zhai, P. (1998). *Get Real: A Philosophical Adventure in Virtual Reality*. Oxford: Rowman and Littlefield.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Personal Space and Design



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The Organization of User Experience

Robert Rosenberger

In 2014, a Google Glass device was strapped to the head of tennis legend Roger Federer as he played a game on the court. Among the Glass's many features, it enables wearers to take pictures or video of whatever stands in their line of vision, activated by a simple tap or swipe across the side of the device, or by voice command. As the smart computing device sat across Federer's face like a pair of eyeglasses, it was set to begin recording. The resulting one minute and twenty second video, shot at Google's facility in Mountain View, provides a striking glimpse of the game from Federer's point of view. It is difficult not to marvel at the intense velocity of play captured in the shaky-cam footage.

Of course, the minor hype that accompanied the video's release made stronger claims. For example, as the Association of Tennis Professionals website put it: "Ever wonder what it feels like to play tennis with Roger Federer? Or even better, as Roger Federer?" (ATP Staff 2014). It might be fair to brush aside such a statement as expected hyperbole, the commonplace exaggeration that accompanies both tech rollout and sports journalism. But it may be useful to interrogate this claim anyway since it might reveal something about the phenomenology of computer interface specifically, and the phenomenology of technology more generally. At the end of this chapter, I'll return to this example with the assistance of some insights into Federer's play offered by David Foster Wallace.

While the Google Glass device failed to make the transition to the mass market, similar technologies are in development. Portable electronics in general are of course utterly commonplace—and mobile and wearable computing are increasingly just as common. Designers strive to create devices that are intuitive and that can fit smoothly into our everyday lives. But how users relate to such devices does not reduce simply to the details of the interface design. With ideas from the phenomenology of technology, we can attempt to conceptualize the nuances of user experience, and how such patterns of experience develop. The particular goal of this chapter is to outline what could be called a 'field theory' of technological experience, one which describes how a user's experience is informed and organized as they use a particular technology, including what stands forward with significance, and what instead goes barely noticed.

I build here on a philosophical perspective called *postphenomenology* (e.g., Ihde 2009; Verbeek 2011; Rosenberger and Verbeek 2015). Founded by American philosopher of technology Don Ihde, postphenomenology combines insights from the phenomenological and pragmatist traditions of thought and applies them to the deep description of concrete human-technology relations. Put another way, this perspective specializes in describing what it's like to use technology. Following, I expand the ideas of postphenomenology to develop a field theory of technological experience.

A central way that postphenomenologists conceive of technology usage is in terms of the notion of 'mediation.' A technology is not simply one of the objects of the world that a person encounters, perceives, or acts upon. A technology comes in between a person and the world,

mediating the relation between the two. It changes the user's encounter with the world, alters their perception, and transforms their capacity to act.

In this way, human-technology relations should be understood to be 'nonneutral.' They do not mediate user experience in an innocent and inactive manner. Technological mediation is actively transformative, potentially enhancing and extending certain aspects of experience, while at the same time potentially limiting or reducing others. We can follow the postphenomenological work of Peter-Paul Verbeek and consider the "co-shaping" role of technological mediation, in which both the user and the world become what they are through the mediation process (2011). The subjectivity of the user and the meaningfulness of the world all exist in the ways they do in a given moment through the mediation of technology.

As a plan for what follows, I would like to distinguish between what could be called the 'form' and the 'organization' of individual human-technology relations. By the form of human-technology relations, I refer to the manner in which the user engages with the device. Do they hold it, or look at it, or speak to it, etc.? Ihde has developed an influential list of these forms, and I review them in the first section. By the organization of human-technology relations, I refer to the particular composition of the various aspects of the user's experience. If we look at the totality of things of which a user is aware at a given moment, then how are those things configured and cast? Which aspects jump forward, and which shrink away? The second section consists of my attempt to develop such a postphenomenological field theory. In a final section I return to the example of tennis and the wearable computing video. With the help of insights from postphenomenology, we can contrast David Foster Wallace's observations about the experience of tennis play with the hype accompanying the Google Glass video. If the video cannot capture Roger Federer's lived experience, then what exactly does it leave out?

The Form of Technologically Mediated Experience

One way that Ihde works to articulate user experience is to distinguish between four different forms of human-technology relations: embodiment relations, hermeneutic relations, alterity relations, and background relations (e.g., 1990). Perhaps more controversially, others have developed the additional notion of 'cyborg relations.' Each of these ideas marks a different manner in which users engage technology, how they approach its interface, how they experience the technology with their bodies, and how the technology transforms their relationship to the world.

To explore these forms of human-technology relations it may be helpful to imagine the experience of someone waking up in the morning and making use of a variety of technologies as they prepare for a typical day.

Embodiment Relations

We can imagine a person waking up to the blaring sound of an alarm clock, almost automatically slapping the largest button to halt the noise, then tossing off a blanket and stumbling out of bed. They step up to a bathroom sink, try to ignore the reflection of the groggy and half-awake individual in the mirror, and with one hand apply toothpaste to a toothbrush held in the other.

As this person begins brushing, an "embodiment relation" is established with the device (Ihde 1990: 73). The toothbrush plays a mediating role, transforming the user's possibilities for action, enabling this person to clean their own mouth. The toothbrush is in a way taken into the user's own bodily experience. That is, as the user is occupied with the task of directing the bristly head of the brush against this or that region of teeth, they come to barely think about gripping the brush in hand. The brush itself has in a way become a transformative extension of

the user's bodily awareness and sphere of possible action. In the vocabulary of postphenomenology, this person "embodies" the device as it is used.

This account of embodiment relations to technology follows a line of thought developed through the classical phenomenological cannon, including, for instance, Merleau-Ponty's example of the woman whose bodily awareness extends through the tall feather on her hat that she navigates beneath low doorways (1945), and Heidegger's example of the shoemaker for whom the work of hammering is more present than the experience of the grip on the hammer used to perform that work (2000). We could of course continue to spell out the embodiment relations involved in the other technological encounters in the story of this typical morning scenario, from the alarm clock button, to the bed and blankets, to the toothpaste tube, to the sink knobs.

Hermeneutic Relations

Imagine next that, perhaps after finishing in the restroom and getting dressed for the day, this person looks toward an analog clock hanging on the wall and realizes it is time for the weather report on the local television news. They pick up a remote control, point it toward the TV across the room, and press the power button. The screen comes to life to reveal a meteorologist standing in front of a large map of the country covered in lines, numbers, and symbols that indicate the movement of weather patterns. Just then, a beeping sound calls from the kitchen to announce that the coffeemaker, preset the night before, has just finished brewing a fresh pot.

Ihde uses the term "hermeneutic relations" to refer to cases in which usage involves perceiving and interpreting a technology's readout (1990: 80). (The word *hermeneutics* otherwise refers to a tradition of philosophy that explores the nature of language meaning and translation.) A technology with which a user shares a hermeneutic relation is one that provides a transformed relation to the world as the user reads the instrumental readout of the device. If a user is already familiar with the practice of interpreting this kind of device, then they may perceive the new information about the world all at once in a perceptual gestalt. A hermeneutic relation is evident in this story when the person looks at and 'reads' the time off of a wall clock display. Since this person is accustomed to the interpretation of this particular readout—that is, since the user already knows how to 'tell time'—they are provided transformed access to the precise time of day in the form a perceptual gestalt. In the case of the analog wall clock, the person does not need to ponder the meaning of each hand and laboriously decipher the clock's meaning; all at once the time of day is announced by the clock face.

A hermeneutic relation to technology can also be seen in the example of the television weather report. The user first establishes an embodiment relation with the remote control, turning on the TV from a distance. But then, looking at the weather map, this person interprets its meaning with regard to the expected weather. As someone already familiar with this act of visual interpretation (through a lifetime of seeing weather maps, and also through watching this particular newscast each morning), they bring a kind of preperceptual readiness to this technological relation, a sort of deeply set perceptual habit that enables an apprehension of the meaning of the weather map all at once.

Then, as the coffeemaker starts beeping from the other room, another hermeneutic relation is established, this time an acoustic one, in which the user listens to and interprets the meaning of the auditory signal.

Where the notion of embodiment relations echoes a line of thought that extends back throughout the phenomenological tradition, Ihde develops the notion of hermeneutic relations and the other forms of human-technology relations reviewed later in this chapter as a critique and expansion of this tradition. He develops these ideas to emphasize that users engage technologies in

a variety of ways, spotlighting just how dependent individual human-technology relations are on their particular practical context. The notion of hermeneutic relations in particular has gone on to find application in a bustling line of postphenomenological work on the topic of imaging technologies, drawing out the roles of human bodily perceptual experience in the practices of scientific and medical imaging (e.g., Ihde 1998; Hasse 2008; Rosenberger 2011a; Rosenberger 2011b; Forss 2012; Wiltse 2014).

Alterity Relations

After watching the weather report and pouring a cup of coffee, this person begins to prepare for the workday. As part of this preparation, they take out a smartphone and activate a personal assistant application, the kind that can be accessed through conversation-style voice interface. This person may ask, “Do I have any meetings scheduled for today?” The device checks through the user’s digital calendar stored on the phone and replies with a prerecorded human-like voice, “You have one meeting at eleven o’clock with your full team.”

Ihde uses the term “alterity relation” to refer to instances of technological mediation in which the interface with the device has a quality similar to that of interacting with another human being (1990: 97). (The word *alterity* has been used in phenomenology to help articulate the special Otherness experienced in the encounter with another person. See esp. Levinas 1969). In the case of an alterity relation to technology, the device takes on a form analogous to interaction with other people. As Ihde puts it, the device takes on a “quasi-otherness” (1990: 98).

An alterity relation is at work in the example of the conversation-like interface of the smartphone personal assistant app, as well as other technologies capable of mimicking human vocal interaction such as dashboard GPS devices that spout off driving directions, or automated telephone customer service voice prompts. Common examples also occur in the form of text interface, such as those found on ATMs or computer screen ‘dialog boxes.’ As our computing technologies continue to advance, and as the field of robotics continues to grow in sophistication, we can expect alterity relations to technology to become more prevalent (e.g., Irwin 2006; Wellner 2014; Bottenberg 2015; Hasse 2015).

Background Relations

Finally this person walks to the front door, ready to leave for the day. But just before turning the knob, they think that it would be a good idea to make certain that the climate control system is on. It may get very hot this afternoon, and this person would like the house to remain a reasonable temperature for the pets. But just before walking back to the thermostat to double check, they suddenly recall hearing the system occasionally clicking on and off quietly all morning. This person is certain they have been hearing it, but this fact has been only barely registered throughout the morning, at least at a conscious level. With everything now in order, this person heads out the front door.

This vague partial awareness of the climate control system is an example of what Ihde calls a “background relation” to technology (1990: 108). In such relations, as the technology mediates a user’s experience of the world, the device itself retains only a quasi-presence in our awareness, remaining “off to the side” (1990: 109). In this way, the technologies to which we share background relations make up the surrounding context of our experience. Such technologies at once transform user experience—protect us from the elements, keep our food from spoiling, heat our water—all while maintaining a partial and backset presence in our awareness, and constituting our technological environment.

Cyborg Relations

As a proposed addition to Ihde's quartet of forms of human-technology relations, some argue that an extra category is needed to account for technologies that are implanted into the human body. It is claimed that such devices cannot simply be included in the category of embodiment relations since they are not merely used in relation with one's body, but are instead 'merged' with the body itself. In everyday cases like pacemakers or stents, or science fiction examples like bionic eyeballs, the device might be understood to blur the line between our conception of the user and the technology. Indeed, such examples challenge the coherence of the very idea of the 'user,' the 'mediation' of technology, and even the notion of 'human-technology' relations. Peter-Paul Verbeek is the chief proponent of this view, and he refers to such scenarios as "cyborg relations" (2011).

In Kirk Besmer's work on this topic, the notion of cyborg relations has been instantiated with the rich example of the cochlear implant, a device used to enable the hearing impaired to experience sound (2012). One part of this technology receives auditory information and sends it as a digital signal to a separate part of the device implanted within a person's head, which then translates that information into a stimulation of the auditory nerve. Besmer argues this to be a clear example of cyborg relations and explores the implications of this device for our conception of human-technology relations.

While it does seem clear that Verbeek and others have identified a crucial issue, the notion of cyborg relations has not proven to be beyond criticism. The trouble is that while a distinction between implanted and nonimplanted devices seems fundamental from a god's-eye-view standpoint, it may not be quite as coherent from a phenomenological perspective (De Preester 2011; Rosenberger 2015b). That is, it is not always apparent what the experiential difference will be between implanted and nonimplanted devices, and if that experience will be distinct and consistent enough to form the basis for a clear conceptual category. From an objective and bird's-eye view of users and technologies, the difference between a technology inside the body and one outside is clear and obvious. But, for example, from the vantage point of the everyday glasses user, those familiar glasses on the face might be much more deeply embodied when compared to a brand new, foreign, and conspicuous implanted device.

The Organization of Technologically Mediated Experience

In addition to considering the form of our relations to technology, it is important to describe what could be called the *organization* of a particular human-technology relation. How can we best describe the user experience of a technology, especially considering that the qualities of that experience may differ between individuals, such as those with different goals, or those with different levels of familiarity with the device? I suggest that what is needed is a 'field theory' of technological experience. We need to find ways to articulate the full experiential field of which a user is aware at a given moment as a device is used, what stands forward, what provides meaningful context, what is barely noticed, and so forth. That is, we need to find ways to describe how technology usage 'reorganizes' a user's 'field of awareness.' (An inspiration here is the work of classical phenomenologist Aron Gurwitsch, 1964, who explored what he called the "organization" of "the field of consciousness").

I have begun developing such a theory over a series of papers, and have thus far considered a technologically informed field of awareness to be potentially characterized by four characteristics (though this list need not be understood as comprehensive): transparency, forefronting, field composition, and sedimentation. By determining when and to what degree these variables

characterize a given case of technology usage, we can begin to capture some of the experiential quality of human-technology relations.

Transparency

A basic phenomenological insight, articulated most famously and in different ways by Martin Heidegger and Maurice Merleau-Ponty, is that as a person makes use of a technology, the device itself withdraws, and the user is instead occupied more with the work being accomplished. Ihde has formulated this idea under the term “transparency” (1990: 73). As a user becomes familiar with a technology, the device takes on a degree of transparency, and the user focuses more on what the technology is used for, and less on the device itself. The default example is a pair of eyeglasses. The glasses mediate user experience through an embodiment relation, transforming a person’s ability to see. For the glasses user who is deeply accustomed to their device, in normal usage it is possible to almost entirely forget they are being worn, all despite the fact that they sit right in the middle of the face and transform the entire field of vision. The earlier account of the remote control is another example of a human-technology relation that can be at times highly transparent. As the user flips through channels on the television across the room, they may grow less aware of the device in hand and the action of pressing buttons as they instead are focused on the changing contents of the television screen.

The notion of transparency is not limited to characterizing embodiment relations, but can pertain across the different forms of human-technology relations. For instance, a deeply transparent hermeneutic relation may involve an immediate interpretation of a technology’s readout. A user’s relation to the wall clock can be deeply transparent. The user does not need to consciously and conspicuously take stock of the clock’s hands and numbers, and explicitly work through what this means with regard to the current time. Instead, the clock’s particular features as well as the user’s own work of interpretation all withdraw, and it is simply the time of day that is acquired at once in a glance.

A rich example can be found in the experience of driving a car. For a driver who is deeply familiar with the act of driving—that is, for someone who ‘knows how to drive’—the many features of the car’s interface take on a considerable degree of transparency. For instance, as the driver operates the vehicle in a normal way, the steering wheel and the pedals withdraw into the background of awareness. A good driver may think actively about their plan to turn at the next intersection, but not so explicitly about the action of turning the steering wheel or applying pressure to the brake. Conversely, a bad driver may pay conscious attention to the relations to the steering wheel and the brake pedal themselves, rather than simply to the tasks of steering and braking.

Even a driver’s sensation of their own body belted into the seat can grow unnoticed. Something similar can be said about the driver’s hermeneutic relationship with the dashboard readouts. For an experienced driver, the apprehension of the information conveyed by dashboard displays takes only a glance. All of this is to say that good driving requires a highly transparent embodiment relation to the car.

Forefronting

As a sort of corollary to the notion of transparency, it is also possible at times to identify features that instead stand forward, taking an explicit position in a user’s overall awareness, and ringing with significance. Let’s refer to this variable as *forefrontedness*. Forefronted features of experience are those that pierce forward to the foreground of the user’s field of awareness. This notion helps to articulate those areas of interest, points of concern, conspicuous objects

of attention, and sounds of alarm that occur within the complexity of technologically mediated experience.

An example can be found in the experience of watching someone give a professional presentation, standing at the front of the room beside Microsoft PowerPoint slides projected onto a large screen, laser pointer in hand. As a member of the audience, many aspects of this experience may take on a considerable level of transparency. Assuming the lecture is good (or at least captivating), as you become engrossed in the performance you may simultaneously grow less aware of the feeling of the chair on which you sit, and less aware of your immediate surroundings, such as the crowd around you and between you and the presenter. Within this technologically mediated experience, certain features at the same time become forefronted. As you focus on the presentation, and look at the screen displaying the PowerPoint slides, the bulleted points featured on those slides stand forward and present themselves as articles of importance. If the presenter also uses the laser pointer, then the dot of the laser on the screen may become forefronted, taking a conspicuous position within your overall field of awareness. Within all the various features of the presentation, some of which may provide different levels of context or may recede into a transparent background almost entirely, some like the bulleted items or laser dot will instead take a forefronted place of prominence.

Field Composition

Another variable of human-technology relations is what could be called *field composition*. This refers to the manner and degree to which the use of a technology reorganizes a user's overall field of awareness. While the notions of forefronting and transparency help to articulate which individual aspects of experience may step forward or back, the notion of field composition points to the way a user's overall field of awareness can become restructured through a user's relation with the technology.

Straightforward examples can be found in devices that reorganize a user's field of vision. As a birdwatcher uses a pair of binoculars, the task of holding the device up to the head, the feeling of the eyepieces against the face, and the action of rolling the focusing wheel, may all take on a degree of transparency. Conversely, when a bird is spotted through the device, it stands forward with significance, forefronted within the context of the user's visual field—and yet a catalog of all the transparent and forefronted aspects of the use of binoculars would not by itself capture the most salient characteristics of the experience. It would not articulate the fact that the user's entire visual experience is transported through the device to another location, separated from their locally situated tactile and auditory sensation, and enframed by a circular ring of darkness to which they may grow deeply accustomed. In this way, we can describe the use of the binoculars as one deeply characterized by field composition. The use of the binoculars involves a substantial recomposition of the user's field of awareness.

An everyday device that significantly recomposes a user's field of awareness focused on auditory experience is the use of the telephone. As a user holds a cell phone beside their face and becomes engrossed in conversation, the device itself may take on a considerable degree of transparency; the forefronted focal point of experience will be the conversation itself, rather than the means of that conversation—the phone. That is, the voice and presence of the interlocutor, and the conversational content, are what draw the user's attention. But something more is happening in the use of the phone. For the user accustomed to talking on the phone, this human-technology relation significantly recomposes a user's overall field of awareness. The interlocutor's voice and the content of the phone conversation stand forward to such a degree that they almost fully occupy the entirety of what the user is aware of in that moment. If, say, standing alone in a room and engrossed in conversation over the phone, the user's awareness

may be so thoroughly configured by this relation that they may not even notice objects on the wall before them, even if this user's eyes remain pointed forward.

In a series of papers I have gone so far as to use this description of the field composition of phone usage as the basis for an account of cell phone-induced driving impairment (e.g., Rosenberger 2012; Rosenberger 2014; Rosenberger 2015a). Indeed it was through this line of research on cell phone driver distraction that I first developed the notion of field composition. My argument is that a driver is habitually inclined to experience the phone in terms of the field composition as described, pulled away from concentration on the road, and drawn instead toward a composition of experience focused mainly on the conversation. This account of the driver distraction of cell phones can be contrasted with the assumptions common to the empirical literature, which largely conceive of this distraction in terms of the inherently limited cognitive resources of the human brain. Both accounts are useful for the effort to raise awareness of the dangers of using the phone while driving, and also of the dangers associated with the continuing development of mobile and wearable computing, dashboard infotainment, and hands-free devices used while behind the wheel.

Sedimentation

The notion of "sedimentation" is used throughout the tradition of phenomenology to refer to the way in which our past experiences remain present to us in a form that provides immediate context to our current experience. The metaphor is to the slow accumulation of detritus that solidifies into sedimentary rock. Our past experiences are present such that our current experience occurs as already meaningful, as already set within the context of significance ordered by the things that have happened to us before. But, as Merleau-Ponty notes, the metaphor to sedimentary rock only extends so far (1945: 131). Where a rock is fixed and inert, the actions of our past experience as a source of immediate significance on our current experience should be understood as active, changing, and formative.

I argue that notion of sedimentation should be adopted within phenomenological discussions of human-technology relations. It should be understood as a variable such that a particular relation could be evaluated as more or less sedimented (e.g., Rosenberger 2012; Rosenberger 2014). In this context, sedimentation can be understood as a kind of meta-variable used to refer to the level of strength of the habits associated with a particular human-technology relation. That is, if a particular human-technology relation can be characterized as one in which certain aspects are transparent, others forefronted, and all this within a certain field composition, then we can also consider just how deeply sedimented this user's relation to this technology is in just these terms.

For example, if a user embodies a pair of eyeglasses with a high level of transparency, then this means that the glasses themselves are not significantly present to the user as they are worn. If we additionally say that the user's relationship with the glasses is highly sedimented, then this transparent relation may pop into place the instant the device is set on the face. As deeply sedimented, the transparency of this relation may resist minor interferences with normal usage, for example possibly remaining deeply transparent despite a smudge on the lens.

The notion of sedimentation plays a crucial role in my account of cell phone driver distraction. I argue that the particular organization of a user's experience associated with normal phone usage, in which the conversation and presence of the interlocutor occupy the user's field of awareness, is one which is also deeply sedimented. It is due to a user's individual long history of using this phone in just this way that their field of awareness is so strongly inclined by habit to take on this particular composition. This has implications for driver safety. Even if a driver is consciously intent on paying attention to the road, I suggest that in moments when the

driving becomes routine or the conversation becomes interesting, they become inclined through habit to take on a field of awareness that is focused mainly on the phone conversation, and not the road.

More on Tennis and Wearable Tech

David Foster Wallace was an author deeply attuned to phenomenology. He didn't hesitate to drop Heidegger's name among those he thought should influence other writers (e.g., Wallace 1997). Much of his famous Kenyon College commencement speech, "This Is Water," reads to me as a straightforward, if inspired, exercise in the phenomenology of everyday life (see 2009, and also across the internet for text and video versions). Wallace's acclaimed *New York Times* article, "Federer as Religious Experience," is a classic in the form of stylized and contemplative journalism, and has been an inspiration for those looking to for those looking to wax philosophical about the game of tennis in general, and about Roger Federer in particular (2006a). It also contains phenomenological insight into what it's like to play at the level of a highly skilled athlete from the athlete's point of view.

Wallace points us to a conversation that took place between Federer and another player, Jonas Bjorkman, whom Federer has just bested decisively in a match. He writes, "Federer and Bjorkman are chatting and joking around, and Bjorkman asks him just how unnaturally big the ball was looking to him out there, and Federer confirms that it was 'like a bowling ball or basketball'" (Wallace 2006a). Wallace breaks down what is being said:

Imagine that you're a person with preternaturally good reflexes and coordination and speed, and that you're playing high-level tennis. Your experience, in play, will not be that you possess phenomenal reflexes and speed; rather, it will seem to you that the tennis ball is quite large and slow-moving, and that you always have plenty of time to hit it. That is, you won't experience anything like the (empirically real) quickness and skill that the live audience, watching tennis balls move so fast they hiss and blur, will attribute to you.

(2006a)

Crucial for us here is the observation that the perception of play from the point of view of an expert athlete while they are in the act of play will be entirely different from the experience of someone watching this play from outside the game.

This difference in perception does not reduce simply to the different things noticed by an expert compared to a nonexpert. Wallace quotes Federer at length on the issue:

for the first time since I have come here to Wimbledon, I went to see a match on Centre Court, and I was also surprised, actually, how fast, you know, the serve is and how fast you have to react to be able to get the ball back, especially when a guy like Mario [Ancic, who's known for his vicious serve] serves, you know? But then once you're on the court yourself, it's totally different, you know, because all you see is the ball, really, and you don't see the speed of the ball.

(Wallace 2006b)

Even for Federer himself, the game looks different from the position of the viewer on the sidelines compared to moments when he himself is playing.

Let's return to the Google Glass video. The earlier quotations, although written years before Glass was invented, seem almost tailor made to trouble the hype claiming that the video supplies the experience of what it's like to be a star player. It is clear that the video shows something

amazing, such as and among other things, Federer's speed. In this way, the video resembles Federer's own experience of watching from the stands, astonished by the velocity of high-level play. The Glass video is notable mainly because this speed is captured from the unusual vantage point of Federer's own face. But as we see in the quotations, the experience of play includes much more.

Some of this experience can be articulated with the terminology developed in this chapter. The description of tennis play can be approached in terms of human-technology relations. For the experienced player, the racket in hand surely becomes embodied to a highly transparent degree in at least many moments of play. Even if in some moments one thinks actively about one's hold on the racket, one cannot spend the bulk of one's conscious attention on the act of gripping the device. The court visually in front of the player, and the zone sensed behind, compose the context of significance. Some aspects of the court are organized through perceptual habituation, attained through years of training, into a field of awareness set with un-ignorable meaning, including the sidelines, the net, and one's own baseline, etc.—and, as Wallace's account attests, for the experienced player the ball itself is forefronted to such a degree that it is described as increasing in size and decreasing in speed. For the experienced player, as the ball accords with sedimented expectations, it is perceived in part in terms of its potential, its readiness to be hit.

These aspects of a player's field of awareness cannot be captured in a simple video, even one filmed from the face of a tennis legend.

References

- ATP Staff. (2014). Tennis Through Federer's Google Glass. May 21, 2014. Available at: www.atpworldtour.com/News/Tennis/2014/05/21/Federer-Edberg-Google-Glass.aspx
- Besmer, K. (2012). Embodying a Translation Technology: The Cochlear Implant and Cyborg Intentionality. *Techné: Research in Philosophy & Technology*, 16(3), 296–316.
- Bottenberg, F. (2015). Searching for Alterity: What Can We Learn From Interviewing Humanoid Robots? In R. Rosenberger and P. P. Verbeek (eds.) *Postphenomenological Investigations: Essays in Human-Technology Relations*. Lexington Books/Rowman Littlefield Press, 175–189.
- De Preester, H. (2011). Technology and the Body: The (Im)Possibilities of Re-Embodiment. *Foundations of Science*, 16, 119–137.
- Forss, A. (2012). Cells and the (Imaginary) Patient: The Multistable Practitioner-Technology-Cell Interface. *Medicine, Health Care & Philosophy*, 15, 295–315.
- Gurwitsch, A. (1964). *The Field of Consciousness*. Pittsburgh, PA: Duquesne University Press.
- Hasse, C. (2008). Postphenomenology: Learning Cultural Perception in Science. *Human Studies*, 31(1), 43–61.
- Hasse, C. (2015). Multistable Robot Ethics. In J. K. B. O. Friis and R. P. Crease (eds.) *Technoscience and Postphenomenology: The Manhattan Papers*. New York: Lexington Books/Rowman Littlefield Press, 169–188.
- Heidegger, M. (2000). *Being and Time*, trans. J. Macquarrie and E. Robinson. Oxford: Blackwell.
- Ihde, D. (1990). *Technology and the Lifeworld*. Bloomington, IN: Indiana University Press.
- Ihde, D. (1998). *Expanding Hermeneutics: Visualism in Science*. Evanston: Northwestern University Press.
- Ihde, D. (2009). *Postphenomenology and Technoscience: The Peking University Lectures*. Albany: SUNY Press.
- Irwin, S. (2006). Technological Other/Quasi Other: Reflection on Lived Experience. *Human Studies*, 28, 453–467.
- Levinas, E. (1969). *Totality and Infinity*. Pittsburgh, PA: Duquesne University Press.
- Merleau-Ponty, M. (1945). *The Phenomenology of Perception*, trans. D. Landes. London: Routledge.

- Rosenberger, R. (2011a). A Case Study in the Applied Philosophy of Imaging: The Synaptic Vesicle Debate. *Science, Technology, & Human Values*, 36(1), 6–32.
- Rosenberger, R. (2011b). A Phenomenology of Image Use in Science: Multistability and the Debate over Martian Gully Deposits. *Techné: Research in Philosophy and Technology*, 15(2), 156–169.
- Rosenberger, R. (2012). Embodied Technology and the Problem of Using the Phone While Driving. *Phenomenology and the Cognitive Sciences*, 11(1), 79–94.
- Rosenberger, R. (2014). The Phenomenological Case for Stricter Regulation of Cell Phones and Driving. *Techné: Research in Philosophy and Technology*, 18(1–2), 20–47.
- Rosenberger, R. (2015a). Driver Distraction From Dashboard and Wearable Interfaces: A Case Against Connectivity. *IEEE Technology & Society Magazine*, 34(4), 88–99.
- Rosenberger, R. (2015b). Postphenomenology: What's New? What's Next? In J. K. B. O. Friis and R. P. Crease (eds.) *Technoscience and Postphenomenology: The Manhattan Papers*. New York: Lexington Books/Roman Littlefield Press, 129–148.
- Rosenberger, R., and P. P. Verbeek (eds.) (2015). *Postphenomenological Investigations: Essays in Human-Technology Relations*. New York: Lexington Books/Rowman Littlefield Press.
- Verbeek, P. P. (2011). *Moralizing Technology*. Chicago: Chicago University Press.
- Wallace, D. F. (1997). Greatly Exaggerated. In *A Supposedly Fun Thing That I'll Never Do Again*. New York: Back Bay Books, 138–145.
- Wallace, D. F. (2006a). Federer as Religious Experience. *New York Times*, August 20, 246. Available at: www.nytimes.com/2006/08/20/sports/playmagazine/20federer.html?pagewanted=all
- Wallace, D. F. (2006b). FOOTNOTES: Federer as Religious Experience. *New York Times*, August 20, 246. Available at: www.nytimes.com/2006/08/20/sports/playmagazine/20federer-footnotes.html?pagewanted=all#footnote8
- Wallace, D. F. (2009). *This Is Water: Some Thoughts, Delivered on a Significant Occasion, About Living a Compassionate Life*. New York: Little, Brown and Company.
- Wellner, G. (2014). The Quasi-Face of the Cell Phone: Rethinking Alterity and Screens. *Human Studies*, 37, 299–316.
- Wiltse, H. (2014). Unpacking Digital Material Mediation. *Techné: Research in Philosophy and Technology*, 18(3), 154–182.

Engineering Design

Pieter E. Vermaas

Introduction

When consulting the design research literature for a characterization of engineering design, two initial issues emerge. The first is that there is in design research a broad and a specific conception of design. The second is that design researchers describe existing practices of engineering design but also develop new practices and prescribe them as (good) engineering design. Both issues have an impact on the question of what engineering design is.

For the broad conception, design researchers may refer to Herbert Simon's characterization that the goal of design is to "devis[e] courses of action aimed at changing existing situations into preferred ones" (1996: 111). Yet, when these same researchers home in on their topic, they may limit engineering design to more specific practices, such as the ones carried out by professional engineers (Lawson and Dorst 2009) or the description of products that solve design problems (Cross 2000). A quick way out of this ambiguity may be that the broad conception lays down what design *in general* is, and the specific one what *engineering* design is. In earlier times, this resolution may have worked, but it currently does injustice to the way engineering design has evolved. For instance, the concept of a product has expanded over the years in engineering (Buchanan 2009) and may now refer to material artifacts as well as to processes, services, innovation strategies, planned actions, and even biological organisms. Resolving the ambiguity by equating engineering design with the broad conception will not do either. It may be argued that in contemporary design all practices that fit Simon's characterization are increasingly carried out. Yet it may be doubted whether all these practices are cases of *engineering* design. For instance, current research in what is called *design thinking* explores how design can realize preferred situations in management, industrial, and societal policy making. But proponents of this design thinking (e.g., Brown 2009; Plattner et al. 2009) advance the view that engineers should participate in such design practices merely on the same footing with professionals from other disciplines, or even subsidiary to them; if such designing is dominated by engineers, so the argument goes, then the technical orientation of engineering would hamper innovation by blocking input from the other disciplines. Hence, engineering design should be taken as broader than the description of products that solve design problems, yet not as too broad since not all design is unconditional engineering design. Laying down *what engineering design is* is therefore linked to fixing what engineering can do.

The second issue implies that a characterization of engineering design is at best provisional. Classic and more recent accounts and surveys of engineering design (e.g., Cross 2000; Kumar 2013) may describe what currently is considered as engineering design. This description may be based on empirical observation with or without theory formation. Yet design research is not just a descriptive discipline that charts existing practices of engineering design; it also prescribes new practices as engineering design, say, ones that are considered to be better than

existing practices, or that broaden what can be achieved with engineering design. These new practices may be derived from existing engineering design (e.g., the account by Pahl and Beitz et al. 2007), be proposed as new practices (e.g., axiomatic design by Suh 2001) or emerge by combining descriptive analysis with new propositions. This possibility of prescribing new practices makes that the characterization of engineering design can change over time in terms of how it is done and, as already noted, in what it may result.

In this chapter, building on my prior work (Vermaas 2015), I survey philosophical issues about engineering design by following how its conception broadened in the last few decades.

Engineering Design for Finding Solutions

By considering design theories and methods originating from the early second half of the twentieth century (e.g., VDI 1993; Pahl and Beitz et al. 2007), engineering design is taken as a practice of finding a technical solution to a design problem. That problem is formulated as a set of physical, technical, and financial requirements that have to be met, and the solution is typically a description of a material artifact. The requirements capturing the problem may need to be articulated by the engineering designers, but the problem itself is fixed by the needs of clients or by technical research targets, and is considered to lie outside of engineering design.

Finding the description of the artifact that solves the problem is typically divided into different phases, such as conceptual design, embodiment design, and detail design (Pahl and Beitz et al. 2007). Engineering design is not a practice that goes linearly through these phases; it may involve iterative steps between phases and subphases, say, when findings in later (sub) phases give reasons to reconsider decisions made in earlier (sub)phases. Conceptual design is aimed at finding a *solution principle* or design concept. This solution principle abstracts from the particulars of the given design problem and is put forward as a solution to the central elements to the problem. This solution is presented as a functional structure, consisting of the overall functions of the artifact to be designed, of the subfunctions that make up that overall functions, and of physical working principles that can realize the subfunctions. More than one solution principle can be generated, which are then evaluated against the requirements capturing the design problem for choosing the best candidate. In embodiment design the overall *material layout* of the artifact is determined, consisting of the shapes, materials, and spatial arrangement of components of the artifact. Again different alternatives may be developed, for exploring possibilities and their (dis)advantages, and these alternatives are again evaluated against the design problem requirements. In detail design the final arrangements, materials, forms, and dimensions of the individual components of the artifact are determined and documented. (Fixing how the artifact is to be manufactured is also part of engineering design and starts in the phase of embodiment design. This aspect of engineering design is not considered here.)

Engineering design, when it proceeds through these phases, is understood as rational reasoning structured by sequences of synthesis, analysis, and evaluation (VDI 1993). In conceptual design and embodiment design, different alternatives are synthesized, a practice that may be taken as more creative. But these alternatives are then systematically analyzed, evaluated, and mutually compared as attempts to meet the requirements capturing the design problem. Engineering design stops when a suitable alternative is found that meets these requirements and for which there are no obvious variations that are technically better or cheaper. Simon (1996) characterizes this pragmatic aspect of engineering design as *satisficing*: in engineering design the space of solutions is not exhaustively searched for finding the globally best solution, since such searches may take up too much time and too many resources; in engineering design one rather settles with finding a locally optimal candidate that meets the set requirements.

Engineering design promotes the use of functions in conceptual design as a means of finding innovative solutions to design problems by, first, formulating a design problem in abstract functional terms, and only then exploring what physical working principles could realize the (sub) functions. Engineers avoid opting immediately for the usual or their preferred overall material layout of the artifact. Still, engineering design is actually often *redesign*, since by considering variations of existing artifacts or of existing components engineers can arrive quickly and cheaply at feasible overall material layouts. Walter Vincenti (1990) calls such redesigning *normal* design in contrast to *radical* design. Yet, radical design is increasingly getting more attention and promotion: the introduction of computers in engineering design has made the often painstaking analysis of solution principles and overall material layouts faster and cheaper, freeing up time and resources for synthesis and evaluation; and innovation in engineering design is increasingly endorsed for economic reasons and for meeting social targets such as sustainability.

Although this description of engineering design as finding technical solutions may be taken as generally accepted, there is a complaint that there is no consolidation in design research of a common conceptual framework and of a coherent understanding of engineering design (e.g., Chakrabarti 2011). Since the early second half of the twentieth century, design research has created an abundance of methods and tools for engineering design, of which some give different descriptions of engineering design as the one given give earlier. This abundance led to a fragmentation of design research, with different strands advancing their own approaches to engineering design and ignoring the findings in other strands (e.g., Blessing and Chakrabarti 2009: 6–8). Even if individual methods may be presented as being state-of-the-art and superseding alternatives, these methods typically coexist side-by-side in the design research literature, leading to survey volumes with names as *101 Design Methods* (Kumar 2013). Moreover, key concepts such as function may have different meanings in different design methods, making comparisons of these methods even more complex.

Contributing to attempts in design research to arrive at a more coherent understanding of engineering design seems a clear task for philosophical reflection on engineering design, as argued and taken up by Per Galle (2008). Analyses to arrive at a common conceptual framework have been carried out specifically for the concept of function (e.g., Houkes and Vermaas 2010). Yet the idea of fixing unambiguous meanings for key concepts as the right course of action may be challenged. Although researchers do not agree about the meaning of, say, the concept of function, consensus in design research is that this coexistence of different meanings should be maintained rather than overcome. The philosophical issue may lay, therefore, not in finding what the precise single meanings of key concepts are, but explaining why they have flexible meanings (Vermaas 2013).

A conceptual issue about engineering design that has been analyzed in detail is how descriptions of artifacts relate intentional and structural descriptions, how in engineering design intentional descriptions—the aim to be realized with the artifact—are translated into structural descriptions of artifacts, as in *The Dual Nature of Technical Artifacts* program (Kroes and Meijers 2006; Kroes 2012). Philosophers in this program observe that intentional and structural descriptions are both needed to capture an artifact, and that these descriptions in that sense complement each other, in contrast to, say, philosophy of mind, where these descriptions can be in competition.

Engineering Design for Finding Problems

The description of engineering design as solution finding may be taken as rationalizing it in two ways: it gives a rational account of the front end of design by assuming that design problems can be captured by sets of requirements; and it presents the process that is to result into

solutions as algorithmic ones in which the analysis and evaluation of candidate solutions proceeds on rational grounds. Descriptive studies of actual design practices led to perspectives on engineering design that are richer and provide more room for creativity. This development is in part motivated by interests in design research in innovation, setting the earlier discussed rational descriptions of engineering design aside as rather prescriptive and (merely) aimed at solutions that count as incremental improvements of existing artifacts.

The scope of engineering design broadened to include the front end of finding design problems with the analysis of design by Donald Schön (1983). Although this work originated in part in architecture, the formulation of the design problem of a client came to be seen as also a part of design engineering. A problem may initially be formulated by the client, but it can be analyzed by the engineering designer not only for articulating it, but also for changing or *reframing* it. Engineering design starts in this analysis with an initial interpretation of the problem by the designer, and with an initial solution or solution direction for finding a product that might solve it. The exploration of this solution direction provides the designer, however, with new insights about the problem. These insights enable the designer to change and improve on the initial interpretation of the problem, thus allowing the designer to choose and explore alternative solutions or solution directions. This reframing of the problem can be seen as a characteristic that sets design apart from regular problem solving in, for example, science and mathematics. Moreover, reframing is sometimes even regarded as necessary in design. Scientific or mathematical problems are said to be well-structured by providing, in the way in which they are formulated, information and success criteria about the kind of solution that is required. Formulations of design problems may in contrast be *ill-structured* (Simon 1984), *wicked* (Rittel and Webber 1984), or *paradoxical* (Dorst 2006). Hence, in design it is sometimes necessary to transform the original formulation of the problem or to transform the design problem itself, in order to make it solvable. This possibility of reframing led to design becoming viewed methodologically as a practice in which the design problem may evolve with the search for its solutions, and that ends when a satisfactory pair is found consisting of a reformulated design problem and a solution to it (Dorst and Cross 2001). The knowledge and ability to, in this way, address ill-structured, wicked, and paradoxical problems of design has been set apart as defining a “third culture” different from the knowledge and abilities to address problems in the other two cultures of the natural sciences and the humanities (Cross 2006).

The phenomenon of reframing design problems is regularly studied in design research in relation to the possibility of arriving at innovation in engineering design, since reframing is seen as an important means to break away from existing perspectives of design research. In design research often a distinction exists between expert designers and novice designers, and analyses focus on the distinctions between the practices and cognitive abilities of these two groups with the aim to understand and to make available to novices the ways experts arrive at telling innovations (e.g., Cross 2006; Visser 2006; Lawson and Dorst 2009), drawing from, for instance, Herbert Dreyfus’s work on expertise. Novices are said to follow a depth-first approach of first exploring alternative design solution directions in detail before selecting one, whereas experts can take a breadth-first approach and select the promising solution directions on the basis of a more coarse-grained assessment of the alternatives. This focus on how experts arrived at solutions in engineering design has in turn led to a new generation of design methods (e.g., Verganti 2009; Dorst 2015) of which a few are mentioned later in this chapter.

Descriptive work on how engineering design is actually done also results in distancing it from the view that design proceeds in a rational algorithmic way. Louis Bucciarelli (1994) describes engineering design as a negotiation process. Engineering design is typically done in teams with members originating from different engineering and nonengineering disciplines (from structural engineering to marketing). These members approach the artifact to be designed

from their own disciplinary perspective, which Bucciarelli calls the disciplinary object worlds of (rational) knowledge and principles. Engineering design is then in part a social process where team members through negotiation and explanation arrive at a (partial) mutual understanding of the artifact as a solution to the design problem. In a more articulated social constructivist approach Wiebe Bijker (Pinch and Bijker 1987; Bijker 1995) has described the development of new artifacts and the decisions made in engineering design for selecting certain solution directions as driven by interactions between social groups—users, clients, industrialists—that are involved in commissioning engineering design or in using the artifacts designed. Latour (1996) describes engineering design as a process in which, symmetrically, the artifacts also drive the process of engineering design, just as the engineers do.

Another challenge to the view that engineering design is primarily a rational practice can be found in more analytic approaches. Maarten Franssen (2005) showed that Arrow's no-go theorem applies to multicriteria analyses and illustrated the problems this causes in engineering design in the evaluation and selection of design solution alternatives. And George Hazelrigg (2003) reviewed the tools used in engineering design for this evaluation and selection—e.g., Quality Function Deployment, Pugh matrices—and demonstrated that these tools can lead to poor decisions and sometimes even to selecting alternatives that are arguably not the best ones.

Presenting engineering design as constituting a third culture amplifies the challenge in articulating the differences between the natural sciences and between the humanities, and in pinpointing the specificity of engineering design. Reframing seems to play a central role in this challenge, and questions that emerge involve what the logical structure of ill-structured, wicked, and paradoxical design problems is, and how reframing addresses them. Calling the answers to the reframed problems 'solutions' to the original problems presupposes such an analysis. First attempts at capturing reframing more formally have been made in design research (e.g., Roozenburg 1993; Dorst 2011), drawing from work by C. S. Peirce by modelling reasoning from problem to solution in engineering design as instances of *abduction*.

The analyses that contest rationality in engineering design raise in turn questions of what standards must be met and how good and bad designs are to be contrasted. The conclusions need not be that engineering design is irrational. Hazelrigg (2003), for instance, concludes his criticism with the proposal that engineering design can incorporate decision theory in evaluating and selecting design alternatives, suggesting that this element of engineering design can be improved to meet the standards of this theory. Also, the analyses that argue for understanding engineering design in terms of interactions between social groups need not necessarily be taken as undermining the value of technological knowledge and rational engineering decisions. As will be discussed later in this chapter, engineering design has broadened to practices that realize values and address societal problems. The social realm thus has become included in engineering design, which allows interpreting these analyzes as making visible the role that social interactions between the various stakeholders play in engineering design. Hence, by stepping over the usual division within philosophy and combining more analytic with more social constructivist works, standards for good engineering design may be within reach.

Engineering Design for Realizing Values

The development of adding the process of problem finding to engineering design implies that engineering design has broadened at its front side. A similar broadening has taken place at the back end of engineering design consisting of the handover of designed products to their manufacturers and their users. Focusing on the latter group, work by Donald Norman (1990) on how users can misinterpret the way products have to be handled, initiated the development

in engineering design to create more intelligible or user-friendly interfaces. The VCR is the classical example of a product that was at one point unintelligible to its users, yet Norman's studies included more simple products such as sliding and revolving doors in buildings, which, despite the good intentions of their designers, created unresolvable problems to the people who wanted to pass through them. Efforts in engineering design to overcome these problems led to *user-centered design* and *interaction design*, again broadening engineering design from a primarily technical problem-solving effort to one in which the expectations and cognitive abilities of users are also addressed. Two quite opposite responses to users can be discerned in this development. First, engineers draw users into engineering design for informing designers about the usability of products, for codetermining how products are designed, and, when applied in a more principled way, for determining what products are designed. The role of the designers then changes increasingly into one of (just) facilitating users in engineering design, as in *participatory design* (e.g., Simonsen and Robertson 2013). The second response is one in which designers continue to take the lead in designing products (as in design thinking, described in the next section) and create usable products on the basis of their expertise and research on users. Extrapolating the point that they can arrive at more useful formulations of design problems than clients, engineering designers are then advanced as experts who can better judge than users what users want and can use, as exemplified by the often-repeated quotes by Henry Ford—"[i]f I had asked people what they wanted, they would have said faster horses"—and Steve Jobs—"[a] lot of times, people don't know what they want until you show it to them."

The focus on usability may be understood as the incorporation of a new value in engineering design. Engineering values such as effectiveness, efficiency, and safety, and commercial economic values, were already a standard part of engineering design, but now engineers are adopting the value of usability also. This development can be taken as part of a wider effort by engineering design to explicitly incorporate in and realize values. In product design, designers no longer approach design as a practice in which needs or wishes of consumers are captured by requirements that in turn define the design problem that is to be solved. Instead designers focus on these users' values, and design directly for users' values, like comfort and well-being. In industrial design, explicit attention to making products meet values such as sustainability, manufacturability, and maintainability that have emerged in design research has resulted in tools and methods for *Design for X*, where X stands for the value focused on. Other clear instances of this value-turn in engineering design concern developments to realize more general societal values, such as in *Design for the Base of the Pyramid* (e.g., Prahalad 2002) aimed at realizing the (basic) needs and values of people in developing countries, and in social design where design is used to address social problems and issues (e.g., Brown and Wyatt 2010). Finally, Batya Friedman (e.g., Friedman et al. 2006), starting in the domain of information and computer technology, developed an explicit *value-sensitive design* approach for enabling the incorporation and support of societal and moral values such as autonomy, trust, and privacy in the products of engineering design.

In philosophy of technology this possibility is taken up as design for moral values (Van den Hoven 2007). Philosophy of technology traditionally criticized the products of technology for their impact on the lives of people and on society as a whole, arguing that these products should realize the moral values that people and society hold (e.g., Winner 1980). A more prescriptive approach emerged with Armin Grunwald's *Technology Assessment* (2009) as an effort to guide the development of technology and its applications in directions that are morally and socially preferred. The proposal to explicitly design for values moves philosophy of technology toward constructively analyzing how specific moral and societal values, ranging from justice to transparency, can be realized in actual engineering design practices, thus enabling responsible innovation in domains from biotechnology to water management (Van den Hoven et al. 2015).

The possibility of designing for values raises philosophical questions about how to operationalize moral and societal values as requirements in engineering design and about how to deal with conflicts between such values (see, e.g., Van de Poel 2009). This possibility confronts another traditional criticism in philosophy of technology: critiques of engineers' ability to determine the goals for which designed products are used. The position that engineers can determine technological use is often discarded as a modernist one that cannot be maintained in the light of empirical studies of how technology develops (e.g., Basalla 1988), often referring to the telephone that was initially designed by Alexander Graham Bell to be an aid to the hard-of-hearing and not for the uses it eventually became to have. Bijker was already mentioned as arguing that social groups besides engineering designers determine the development, use, and meaning of the designed products. Don Ihde (2008) argues that it is an intentional fallacy to hold that the intentions of engineering designers determine the uses of products, or with design for values, fix the moral and societal values these products realize. Peter-Paul Verbeek (2005; 2011) takes a middle position in this criticism of engineering design by taking products when used as *mediating* the experience and lives of users, leading to in interplay between and codevelopment of the original uses for which the products are designed and the uses and meanings these products eventually get for users. With this middle position, effective design for values should in some way incorporate the dynamics of this mediation.

Design Thinking

The developments sketched earlier have served to broaden the scope of engineering design considerably. Where in the beginning of the second half of the twentieth century, engineering design involved finding material artifacts as technical solutions to fixed design problems, design research enriched engineering design by descriptive work charting existing practices and by prescriptive work adding new practices. Engineering design developed in this way into a practice in which designers can interpret and reframe problems, in which users and other stakeholders can be considered and actively drawn in, and in which the resulting solutions can be not only material artifacts but also processes, services, innovation strategies, planned action, and organisms—and where originally design solutions should be effective, efficient, and safe. Engineering design can now explicitly be aimed at realizing other moral and societal values, such as usability, sustainability, and justice. Moreover, these design practices are currently under the heading of *design thinking* and are no longer applied to only the traditional engineering domains of product and utility development. Design is, for instance, applied to the social and public domains for addressing societal and urban policy issues that so far have been left to politics (e.g., Brown and Wyatt 2010; Dorst 2015), and to the commercial domain for creating innovative products for firms, as well as management tools for transforming those firms into more dynamic organizations (e.g., Brown 2009; Plattner et al. 2009; Verganti 2009).

In design thinking, the expertise of designers to interpret and reframe problems and the potential of design practices to arrive at innovation is further emphasized. Design thinking practices include front-end phases in which designers actively research their clients and prospective users to determine and better articulate their needs and values, using ethnographic studies and empathy (e.g., Brown 2009; DSchool 2011). Designers may even start projects autonomously by studying people to find ways to improve their lives, skipping clients in favor of providing a first formulation of the problem or for commissioning projects. Design thinking furthermore aims at innovative results by drawing in the expertise of different professions and different stakeholders. Teams engaged in design thinking should moreover be genuinely multidisciplinary rather than be dominated by one discipline to avoid that the search for solution directions gets limited to those that are endorsed within the dominating discipline (e.g., Brown

2009; Plattner et al. 2009). Collaboration with parties other than clients and users is stimulated to draw in people who embody or can identify new cultural trends (e.g., Verganti 2009), or to include organizations and other stakeholders involved in the design problem or solution directions (e.g., Dorst 2015). As such there are similarities with participatory design, yet, in design thinking, the designers remain in control: they create an understanding of the problems, needs, and values of users; they draw in the expertise of other people and disciplines, analyzing the values of stakeholders; and they eventually add their own perspective, vision, or values to frame the design problem and explore solution directions (e.g., Plattner et al. 2009; DSchool 2011; Dorst 2015). Design thinking may therefore be said to be propositional: designers do not derive a solution to a problem, but on the basis of their expertise and analyses propose new products, following Jobs's maxim that people do not know what they want until designers show it to them.

With the emergence of design thinking it seems that engineering design grew into Simon's broad conception of devising courses of action aimed at changing existing situations into preferred ones. Yet the term used is *design thinking* and not *engineering design thinking* or *engineering thinking*. Conceptually there seems to be no barrier to taking design thinking as engineering design, and in engineering research, though small, there has been a tradition that advances and explores this broad conception of engineering—as say Simon (1996), Hubka and Eder (1988)—systems engineering, and contemporary interests in designing sociotechnical systems (e.g., Bauer and Herder 2009). A reason to nevertheless set engineering design apart is the emphasis that design thinking should be conducted by multidisciplinary teams. Design thinking presents itself as a means to innovate by leading to solutions that were not yet found by traditional means, and in this way takes engineering design as conservative: designing engineers then still have a place in design thinking, but should contribute in collaboration with professionals from other disciplines, on the same footing with or even subsidiary to them.

Whether engineering design will eventually be equated with design thinking may be largely a sociological issue concerning the professional identity of engineers. Yet if the distinction remains, it poses the tasks of elucidating both design thinking and engineering design, and of understanding their relations (see, e.g., Downey 2005; Christensen et al. 2015).

Independently of this distinction, design thinking can be taken as advancing most clearly the problem finding and solving aspects of engineering design. For analyzing the claim that engineering design represents a third culture, it therefore seems apt to focus on design thinking for analyzing the specificity of engineering design and for giving the relation between designing and problem solving in the natural sciences and humanities.

With its ambitions to generate innovative solutions to problems in a broad spectrum of domains, design thinking becomes even more subject to the critical tradition in philosophy of technology that challenges the modernist idea that designers can determine the uses and meanings of their solutions. A parallel may be drawn with earlier ambitions in engineering design to resolve or alter social problems by technological interventions. After this possibility was introduced under the heading of *technological fix* (Weinberg 1966), the term later started to represent the naivety involved in the engineering assumption that technology can conclusively address social issues and that engineers can control all the side effects of the use of new technological applications. On the one hand, one can argue that the way in which design thinking arrives at its solutions is more sophisticated than how engineers created their technological fixes, since design thinking explicitly includes users, clients, and other stakeholders, and considers their moral and societal values. Yet, on the other hand, the solutions that design thinking produces are supposed to be innovative and be 'game-changers' that, like Jobs's product innovations, are supposed to disrupt existing situations (think of how the Apple iPod changed the music CD market). It may therefore also be argued that design thinking eventually will be confronted

with unanticipated side effects as well, a phenomenon that design thinking actually seems to welcome.

Design Science

Design research has produced ample material to warrant and enable philosophical analyses of engineering design. These analyses will be confronted with the fragmentation of design research in different strands and with the lack of consolidation of its conceptual framework and findings. Philosophy may contribute to this consolidation and will then face a final and in design research controversial issue of whether it should aim at a science of design.

Design itself is not a scientific activity, just as the behavior of elementary particles and the actions of human beings in social contexts are not. Yet, as the disciplines of physics and sociology are scientific ones, also design research can be so. But this possibility is in design research not generally endorsed. Design research has produced work aimed at general theoretical frameworks for capturing engineering design, as for instance Simon's (1996) sciences of the artificial, Hubka and Eder's (1988) theory of technical systems, Suh's (2001) axiomatic account of engineering design, and more recently the C-K theory of design by Hatchuel et al. (2013). Design research has known periods in which more visionary perspectives were developed (Buchanan 2009). Yet, in mainstream design research such general design theories seem to play a minor role; rather than launching efforts to determine which of these theories are most accurate or useful, or how they can be improved, design researchers seem to create a multitude of accounts of and methods for engineering design which are accepted to coexist side-by-side, creating a toolbox from which designer can draw at will (e.g., Kumar 2013). This does not mean that design research is altogether unscientific or untheoretical. In descriptive research on engineering design processes, design researchers import the scientific practices of statistics, psychology, and the cognitive sciences. In addition, there are efforts to arrive at research methodology to validate the prescriptive claims of methods and tools for engineering design (e.g., Seepersad et al. 2006; Blessing and Chakrabarti 2009; Vermaas 2014) in which also research methodology for validating prescriptive claims in medical research is considered (Frey and Dym 2006). Yet an active effort to critically test and compare the different engineering design methods and tools is largely absent in design research—and design researchers again take distance from the established sciences by advancing that the explorations done in design practices can itself be taken as a (new) kind of research (known as, e.g., *research through design*) that does not comply with the current standards of scientific research (e.g., Frayling 1993; Koskinen et al. 2011).

The controversy about aiming at a science of design may be related to the third culture claim. As the first culture of the natural sciences created standards that made it problematic for the humanities to establish themselves as scientific, also design research is confronted with a struggle with positioning itself as a science relative to the standards that are now set by the natural sciences and humanities. Yet with the sketched broadening of engineering design to an increasingly general problem solving approach applicable in a wide range of domains, having well-articulated theories and research methodology becomes equally increasingly important. With this broadening, engineering design is no longer a niche practice that produces on demand material artifacts, but becomes one by which we all—engineers, scientists, industrialists, policy makers—shape and innovate our material and social reality and realize our (moral) values in it. An understanding of what engineering design is and of how the methods and tools for doing it can be validated, becomes then central to its effective application and dissemination. Philosophical research may provide a major contribution to this understanding, constructively by

articulating its conceptual and scientific basis, and critically by challenging the modernist claim that engineering design will solve our problems.

References

- Basalla, G. (1988). *The Evolution of Technology*. Cambridge: Cambridge University Press.
- Bauer, J. M., and Herder, P. M. (2009). Designing Socio-Technical Systems. In A. W. M. Meijers (ed.) *Philosophy of Technology and Engineering Sciences*. Amsterdam: Elsevier, 601–630.
- Bijker, W. E. (1995). *Of Bicycles, Bakelites and Bulbs: Towards a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press.
- Blessing, L. T. M., and Chakrabarti, A. (2009). *DRM: A Design Research Methodology*. London: Springer.
- Brown, T. (2009). *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*. New York: Harper Business.
- Brown, T., and Wyatt, J. (2010). Design Thinking for Social Innovation. In *Stanford Social Innovation Review*, Winter 2010 edition, 30–35. Available at: www.ssireview.org/articles/entry/design_thinking_for_social_innovation/ [Accessed April 5, 2015]
- Bucciarelli, L. L. (1994). *Designing Engineers*. Cambridge, MA: MIT Press.
- Buchanan, R. (2009). Thinking About Design: An Historical Perspective, in Engineering. In A. W. M. Meijers (ed.) *Philosophy of Technology and Engineering Sciences*. Amsterdam: Elsevier, 409–453.
- Chakrabarti, A. (2011). Towards a Taxonomy of Design Research Areas. In H. Birkhofer (ed.) *The Future of Design Methodology*. Springer: London, 249–259.
- Christensen, S. H., Didier, C., Jamison, A., Meganck, M., Mitcham, C., and Newberry, B. (eds.) (2015). *Engineering Identities, Epistemologies and Values: Engineering Education and Practice in Context, Volume 2*. Dordrecht: Springer.
- Cross, N. (2000). *Engineering Design Methods: Strategies for Product Design*, 3rd edition. Chichester: John Wiley & Sons.
- Cross, N. (2006). *Designersly Ways of Knowing*. London: Springer.
- D.School (2011). *D.School Bootcamp Bootleg*. Available at: <http://dschool.stanford.edu/wp-content/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf> [Accessed April 4, 2015]
- Dorst, K. (2006). Design Problems and Design Paradoxes. *Design Issues*, 22(3), 4–17.
- Dorst, K. (2011). The Core of ‘Design Thinking’ and Its Application. *Design Studies*, 32, 521–532.
- Dorst, K. (2015). *Frame Innovation: Create New Thinking by Design*. Cambridge, MA: MIT Press.
- Dorst, K., and Cross, N. (2001). Creativity in the Design Process: Co-Evolution of the Problem-Solution. *Design Studies*, 22, 425–437.
- Downey, G. (2005). Are Engineers Losing Control Over Technology: From ‘Problem Solving’ to ‘Problem Definition and Solution’ in Engineering Education. *Chemical Engineering Research and Design*, 83(A6), 583–595.
- Franssen, M. (2005). Arrow’s Theorem, Multi-Criteria Decision Problems and Multi-Attribute Design Problems in Engineering Design. *Research in Engineering Design*, 16, 42–56.
- Frayling, C. (1993). Research in Art and Design. *Royal College of Art Research Papers*, 1, 1–5.
- Frey, D. D., and Dym, C. L. (2006). Validation of Design Methods: Lessons From Medicine. *Research in Engineering Design*, 17, 45–57.
- Friedman, B., Kahn, Jr., P. H., and Borning, A. (2006). Value Sensitive Design and Information Systems. In P. Zhang and D. Galletta (eds.) *Human-Computer Interaction in Management Information Systems*. Armonk, NY: M.E. Sharpe, 348–372.
- Galle, P. (2008). Candidate Worldviews for Design Theory. *Design Studies*, 29, 267–303.
- Grunwald, A. (2009). Technology Assessment: Concepts and Methods. In A. W. M. Meijers (ed.) *Philosophy of Technology and Engineering Sciences*. Amsterdam: Elsevier, 1103–1146.
- Hatchuel, A., Weil, B., and Le Masson, P. (2013). Towards an Ontology of Design: Lessons From C-K Design Theory and Forcing. *Research in Engineering Design*, 24, 147–163.
- Hazelrigg, G. A. (2003). Validation of Engineering Design Alternative Selection Methods. *Engineering Optimization*, 35, 103–120.

- Houkes, W., and Vermaas, P. E. (2010). *Technical Functions: On the Use and Design of Artefacts*. Dordrecht: Springer.
- Hubka, V., and Eder, W. E. (1988). *Theory of Technical Systems*. Berlin: Springer.
- Ihde, D. (2008). The Designer Fallacy and Technological Imagination. In P. E. Vermaas, P. Kroes, A. Light, and S. A. Moore (eds.) *Philosophy and Design: From Engineering to Architecture*. Dordrecht: Springer, 51–59.
- Koskinen, I., Zimmerman, J., Binder, T., Redström, J., and Wensveen, S. (2011). *Design Research Through Practice: From the Lab, Field, and Showroom*. Waltham, MA: Morgan Kaufmann.
- Kroes, P. A. (2012). *Technical Artefacts: Creations of Mind and Matter—A Philosophy of Engineering Design*. Dordrecht: Springer.
- Kroes, P. A., and Meijers, A. W. M. (2006). The Dual Nature of Technical Artefacts. *Studies in History and Philosophy of Science*, 37, 1–4.
- Kumar, V. (2013). *101 Design Methods: A Structured Approach for Driving Innovation in Your Organization*. Hoboken, NJ: Wiley.
- Latour, B. (1996). *ARAMIS or the Love of Technology*. Cambridge, MA: Harvard University Press.
- Lawson, B., and Dorst, K. (2009). *Design Expertise*. Oxford: Architectural Press.
- Norman, D. A. (1990). *The Design of Everyday Things*. New York: Perseus.
- Pahl, G., Beitz, W., Feldhusen, J., and Grote, K.-H. (2007). *Engineering Design: A Systematic Approach*, 3rd edition. London: Springer.
- Pinch, T. J., and Bijker, W. E. (1987). The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. In W. E. Bijker, T. P. Hughes, and T. J. Pinch (eds.) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT, 17–50.
- Plattner, H., Meinel, C., and Weinberg, U. (2009). *Design Thinking: Innovation Lernen—Ideenwelten Öffnen*. Munich: mi-Wirtschaftsbuch.
- Prahalad, C. K. (2002). *The Fortune at the Bottom of the Pyramid*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Rittel, H. W. J., and Webber, M. M. (1984). Planning Problems are Wicked Problems. In N. Cross (ed.) *Developments in Design Methodology*. Chichester: Wiley, 135–144.
- Roozenburg, N. F. M. (1993). On the Pattern of Reasoning in Innovative Design. *Design Studies*, 14(1), 4–18.
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. London: Temple Smith.
- Seepersad, C. C., Pedersen, K., Emblemavåg, J., Bailey, R., Allen, J. K., and Mistree, F. (2006). The Validation Square: How Does One Verify and Validate a Design Method? In K. E. Lewis, W. Chen, and L. C. Schmidt (eds.) *Decision Making in Engineering Design*. New York: ASME, 303–314.
- Simon, H. A. (1984). The Structure of Ill-Structured Problems. In N. Cross (ed.) *Developments in Design Methodology*. Chichester: Wiley, 145–166.
- Simon, H. A. (1996). *The Sciences of the Artificial*, 3rd edition. Cambridge, MA: MIT Press.
- Simonsen, J., and Robertson, T. (eds.) (2013). *Routledge International Handbook of Participatory Design*. New York: Routledge.
- Suh, N. P. (2001). *Axiomatic Design: Advances and Applications*. New York: Oxford University Press.
- Van de Poel, I. (2009). Values in Engineering Design. In A. W. M. Meijers (ed.) *Philosophy of Technology and Engineering Sciences*. Amsterdam: Elsevier, 973–1006.
- Van den Hoven, J. (2007). ICT and Value Sensitive Design. In P. Goujon, S. Lavelle, P. Duquenoy, K. Kimppa, and V. Laurent (eds.) *The Information Society: Innovation, Legitimacy, Ethics and Democracy: In honor of Professor Jacques Berleur s.j.*. Boston, MA: Springer, 67–72.
- Van den Hoven, J., Vermaas, P. E., and Van de Poel, I. (eds.) (2015). *Handbook of Ethics, Values, and Technological Design*. Dordrecht: Springer. Available at: <http://link.springer.com/referencework/10.1007/978-94-007-6970-0>
- VDI (1993). *VDI Guideline 2221: Methodik zum Entwickeln und Konstruieren technischer Systeme und Produkte*. Düsseldorf: Verein Deutscher Ingenieure.
- Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Pennsylvania State University Press.

- Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press.
- Verganti, R. (2009). *Design Driven Innovation: Changing the Rules of Competition by Radically Innovating What Things Mean*. Boston, MA: Harvard Business Press.
- Vermaas, P. E. (2013). On the Coexistence of Engineering Meanings of Function: Four Responses and Their Methodological Implications. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 27, 191–202.
- Vermaas, P. E. (2014). Design Theories, Models and Their Testing: On the Scientific Status of Design Research. In A. Chakrabarti and L. T. M. Blessing (eds.) *An Anthology of Theories and Models of Design: Philosophy, Approaches and Empirical Explorations*. London: Springer, 47–66.
- Vermaas, P. E. (2015). Design Methodology and Engineering Design: From Technical Problem Solving to Social Exploration. In S. H. Christensen, C. Didier, A. Jamison, M. Meganck, C. Mitcham, and B. Newberry (eds.) *Engineering Identities, Epistemologies and Values: Engineering Education and Practice in Context, Volume 2*. Dordrecht: Springer, 147–159.
- Vincenti, W. G. (1990). *What Engineers Know and How They Know It: Analytical Studies From Aeronautical History*. Baltimore, MD: John Hopkins University Press.
- Visser, W. (2006). *The Cognitive Artifacts of Designing*. Mahwah, NJ: Lawrence Erlbaum.
- Weinberg, A. M. (1966). Can Technology Replace Social Engineering? *University of Chicago Magazine*, 59, 6–10.
- Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109, 121–136.

Further Readings

- Birkhofer, H. (ed.) (2011). *The Future of Design Methodology*. London: Springer. (A volume with contributions reviewing the current state of research on engineering design as finding solutions.)
- Buchanan, R. (2009). Thinking About Design: An Historical Perspective, in Engineering. In A. W. M. Meijers (ed.) *Philosophy of Technology and Engineering Sciences*. Amsterdam: Elsevier, 409–453. (A description of the developments in the thinking about design in the twentieth century tracing their relations with philosophy.)
- Galle, P. (ed.) (2002). *Design Studies*, 23(3), 211–361. (A special issue bringing together work in philosophy of design.)
- Lawson, B., and Dorst, K. (2009). *Design Expertise*. Oxford: Architectural Press. (An accessible account of design that integrates current results of design research.)

Design

Ann Johnson

Design is an important element in the philosophy of technology. It is both an activity and process (a verb), and an outcome (a noun). Designers design designs—this is a legitimate sentence, but one that offers no clarity at all about the nature of either the activity of designing or what constitutes a design. Is a design a plan or representation (for example, a blueprint) or should we consider that a design is an artifact, or a physical and material thing? In everyday language both of these uses are accurate and common—a design is a plan and an artifact is designed. One might offer a sequential account of designing: the activity produces both the plan and the artifact, in most cases in that order (Ferguson 1994). And the execution of the design requires consideration of the physical form that the artifact will take once it is made; making then becomes an input to the process as well as an outcome. But designing is interesting philosophically in large part because it does involve the execution of a plan and produces an artifact—designs merely on paper (for example, Leonardo DaVinci’s flying machines) have garnered little attention as designs, because they’re only plans and not artifacts (that, for example, DaVinci was able to build, test, and use). Thus, philosophically, most of the interest is directed toward technical artifacts, because the combination of their material nature with the knowledge-producing activity of designing is what’s at the heart of philosophy of engineering, and to a lesser extent a legitimate question in the philosophy of science. For this reason nonmaterial or nonphysical designs, such as graphic designs, are also less well-examined, although they do share the combination of both functionality (they do something) and a particular, designed form. It’s not to say that graphic design isn’t worthy of study, but rather than in the philosophy of technology the materiality of technology is a challenging problem that has attracted more attention than the aesthetics of design, at least among philosophers of technology and especially philosophy of engineering. As a result, the subject of this chapter will be largely the design of material technologies and technical artifacts, and the designers under consideration will primarily be engineers.

That said, it would be misleading to think of engineers as the only or even the most exemplary kind of designers. In some ways, engineering design is exceptional and works differently than other communities’ and professional groups’ efforts at design. This has to do with the attention to the highly technical functions and complex physical forms of the kinds of things engineers design, like machines. Designers outside engineering also grapple with—the convergence of form and function; this is most obvious in the case of architecture, where Louis Sullivan famously wrote “form ever follows function” in “The Tall Office Building Artistically Considered,” an article in Lippincott’s magazine in 1896. Sullivan and subsequent modernist architects embraced and clarified this idea, while their critics railed against it as inhuman, uncreative, and boring, or degenerate. The phrase “form follows function” has had similarly polar reactions in other fields but is a good marker of the dual nature of designed objects—they

do have both a function and a physical form and design is the art and science of articulating the relationship in a material reality (deRidder 2007).

Form and Function in Design

Despite the focus here on engineers, designers come from many different professional communities. Architects are a notable community of designers, as are industrial designers. Industrial designers are at times difficult to distinguish from engineers, but they tend to focus on the aesthetics or the form, while engineers tend to concentrate on the function and mechanics. Yet the rhetorically powerful distinction between form and function isn't so clear in an actual artifact. For example, famous industrial designer Norman Bel Geddes designed the case for the Harvard Mark I computer (sometimes also referred to as the IBM Automatic Sequence Controlled Calculator)—this case comprises the outward appearance of the machine, which was completed in 1944. Yet Bel Geddes had to work to specification; his contribution was constrained by the electrical engineering inside the case. What was inside was sketched out by physicist Howard Aiken and then executed by a team of IBM engineers. Who designed the Mark I? Everyone involved did—they all contributed to the function and the form of the technical artifact. This adds an additional layer of complexity to the philosophical study of design: it is a social, as well as epistemological and aesthetic, activity. All three of these dimensions must stay under consideration when analyzing designs. The social applies primarily to the activity, while the epistemic and aesthetic are components of both the activity and the outcome.

The complexity of design as a category of activity and outcome results from this triad of concerns. Designs have what Peter Kroes has called “a dual nature of structure and function” and are socially produced (Kroes 2000, Meijer 2000, de Ridder 2007). Producing the artifact that entangles concerns about its functionality and its unique and material structure requires that designers make new knowledge. Therefore, design also constitutes an epistemic activity, wherein new knowledge is an outcome. Sometimes the knowledge produced in the process of designing an artifact is of a technical and/or scientific propositional form (say, a new algorithm for predicting an artifact's behavior which may be essential to designing a functioning artifact). Not all designs produce such propositional knowledge, but it is not unusual. This sort of knowledge is a very common object of philosophical analysis, especially in the philosophy of science. There are active debates about the nature of engineering knowledge and whether or not it is distinctive and/or derivative from scientific knowledge (Kroes 2000, Layton 1971, Pitt 2000). It is important not to restrict or reduce this new knowledge to being merely an application of what is already known, because the new knowledge produced isn't (necessarily) about nature or its laws, but rather about the artifact. Since the artifact is new, the knowledge about it is new.

Design Epistemology

Designed artifacts pose a more opaque epistemological question: to what extent should we consider the artifact itself as a material, nonpropositional form of knowledge? A philosophy of design must attend to the materiality of technology, as well as its more philosophically accessible, traditional, epistemic aspects. Epistemology is more comfortable territory for philosophy; tools, frameworks, and theories abound to explain the nature of propositional knowledge. The nature of materiality is poorly studied in comparison. But designed artifacts complicate even the materiality of things by combining their material and epistemological aspects. When knowledge is borne by material objects, it raises the question of how things can constitute a category of knowledge, one very different from the verbal, propositional knowledge that is the common

focus of philosophical epistemology. Davis Baird's *Thing Knowledge* focuses on the materiality of knowledge in a specific category of designed devices: scientific instruments. Baird (2004) argues that we ought to understand these artifacts as a type of knowledge, not merely the outcome of knowledge production. In looking at Michael Faraday's electric motor, Baird urges us to understand that knowledge resides in the action of the instrument; that Faraday is unable to articulate the phenomenon without the instrument itself. He writes that the motor "bear[s] knowledge of a kind of material agency" (Baird 2004: 15). The motor developed in partial independence of Faraday's theory; it cannot be reduced to an instance of theory and bears knowledge materially.

Failure in Design

An important, but perhaps strange, element of engineering design is the process of learning from failures. Engineering designs almost always fail under some set of conditions, and the reasons for failures can be very epistemically important. Bridges, as a spectacular example, fail due to many aspects: loading, strength of materials, misunderstandings of key geological features, poor modeling, and even mathematical errors (Petroski 1992; Petroski 1999; Rosenberg and Vincenti 1978; Kranakis 1996). Understanding the conditions under which a technology collapses or simply fails to function is a critical element in producing knowledge for design, as well as an important step in producing an improved design. Engineering failures need robust institutional, as well as informal, structures to allow knowledge of failure to feed into communities of designers. An example of this is the effect of the collapse of the Tacoma Narrows Bridge. Bridge designer Othmar Ammann wrote in 1953

the Tacoma Narrows bridge failure has given us invaluable information. . . . It has shown that every new structure which projects into new fields of magnitude involves new problems for the solution of which neither theory nor practical experience furnish a practical guide. It is then that we must rely largely on judgement and if, as a result, errors or failures occur, we must accept them as a price for human progress.

(Billington 1983: 37)

Social Aspects of Design

In producing knowledge, whether of a traditional or material form, designers work in groups most of the time. While there are examples of lone geniuses working in isolation, these stories often prove to be far more constructed for rhetorical purpose than accurate accounts. Even Edison, with 2,332 patents in his own name, was engaged in a highly social activity, creating a laboratory in Menlo Park, New Jersey first, then later a larger laboratory in West Orange, New Jersey to house the growing number of collaborators in close proximity to one another who would industrialize the production of new designs. Modern technologies are most commonly produced in industrial and corporate settings, or sometimes in collaborations among academic engineers. Louis L. Bucciarelli writes about the complex social relations that develop around system design. In his book *Designing Engineers*, Bucciarelli (1994) makes the point that no single individual understands how the complete design of the telephone system works. Each designer understood the system from the point of view of the component on which they were working and about which they possessed expert knowledge. They saw the other components only through this perspective. He writes, "inside each firm, there are different interests, perspectives, and responsibilities—corporate planning, engineering, research, production, marketing, servicing, managing—and consequently different ways in which the telephone 'works'"

(Bucciarelli 1994: 3) Bucciarelli describes these different viewpoints as “object worlds,” wherein different visions of technical artifacts in a system are thought of in fundamentally different ways. Designing them requires these multiple perspectives. Thomas Parke Hughes refers to the people, usually engineers, who assemble these systems as “system builders” and includes far more than just the technical artifacts in the systems, also taking account of legal structures, financial arrangements, natural and human resources, and so on in the description of what it means to design electrical power distribution grids (Hughes 1987: 51–52). Design, like engineering more generally, is a social activity, and there are many different arrangements of people, artifacts, and knowledge (Constant 1987).

Collaboration occurs among engineers working directly together on a design, but design is also a social activity within a community of practitioners who might well be working on competing designs. These problem-oriented knowledge communities are all producing knowledge in the endeavor (Johnson 2009). The community is focused on solving a particular problem—that is, determining what the function of the technology should be, then how to design an artifact to fulfill that function. The community, though, is dynamic in its membership, and new participants will come in and out of the community, depending on the shape the artifact is taking and how its function may be changing. I will offer a concrete example of this process next. Thinking of design as a socioepistemological activity then means that the community of practitioners who produce designs (often multiple competing designs) is co-constructed with the knowledge needed to produce designs. Neither knowledge nor community is established in a complete and stable form before the other. This places the activity of socioepistemic design on two levels: (1) collaboration over a design and (2) communal in forming a community of practitioners, who may be working for competing firms. Different kinds of knowledge will be produced and used in different ways at the two levels of social activity in design.

A Case Study of Design: The Case of Antiskid Devices

In addition, designs are also functional, and this functionality has epistemic dimensions: How do we know how to make a device that fulfills the function? How do we know what the function itself is? Decisions about this can occur both with a design team and among a community of practitioners. The more complex and surprising negotiation occurs among a community of practitioners. Let me offer an extended example, drawn from my book examining the production of knowledge in the process of design, *Hitting the Brakes: Engineering Design and the Production of Knowledge* (2009), which describes the design of antilock braking systems for automobiles over the period ~1950 to the mid-1970s.

Following World War II, government officials in Great Britain began to consider road safety. This occurred in Britain not necessarily because roads were less safe in Britain than other nations, but rather because Britain did a better job of keeping nationalized road statistics. The statistics showed that cars crashed because drivers lost control of their cars while they were stopping. This was the problem around which the initial community formed—professionals interested in the problem of car crashes. This early community included government officials, police officers, statisticians, road designers, tire designers, and automotive engineers. For several years, officials at public agencies in Europe and the United States designed instruments to measure cars’ performances on dry roads, wet roads, snow and ice, curves and straight roads, paint, and different compositions of road materials from asphalts to concretes to gravel. They then used the results from tests of vehicles designed for testing and test results produced by attaching measuring instrument to vehicles in use on the roads. These test results allowed the knowledge community to collectively refine their definition of the problem of skidding and simultaneously grow the community of researchers focusing on skidding. The problem-oriented

knowledge community was forming, and designs of instruments, tests, and statistical models were the knowledge products. All were asking the question: what aspects of cars, drivers, and roads created the phenomenon of skidding?

The first stage of refining the question of skidding hinged on creating a better understanding of the interaction between the car, driver, and road. New designs for cars' steering and braking systems, and their tires, or redesigning road surfaces, markings, and layouts, depended on knowledge inputs that pointed at concrete components to improve and in what particular ways. Designs typically improve iteratively, based on new information about the performance of the design under examination. These iterations are often represented by flowcharts with varying levels of detail showing cycles of design development, beginning with problem identification, research, conceptualization and specification, design production, testing, analysis and looping back to research and conceptualization for the next round (Hill 1970: 36). In engineering design textbooks, the knowledge production aspect is typically underemphasized, often subsumed under the terms "research" or "analysis." In fact, in making new designs, engineers spend much time producing models and explanations of a prototype's behavior—both its positive and negative aspects. In the case of anti-skid devices, early problem definition and research focused on sorting out the contributions various components of cars and roads made to skidding so that the problem would be narrowed from skidding writ large to skidding caused by poorly functioning braking systems. As the community defined the problem as one produced by braking systems, it also narrowed the focus for future research. Engineers and road researchers decided that skidding would be defined as what happened when the brakes locked up on the wheel, that is, when the friction between the brake and wheel exceeded the friction between the tire and road and the wheel just slid across the road, instead of rolling. This was a contingent definition, and they could have defined skidding differently. But in constructing this working definition they specified a direction that designs would take, what functions they would produce, and what would be considered a solution to the problem of skidding. This orientation also helped define who would have something to contribute to the process of design and be a part of the knowledge community.

After designing several classes of new instruments to measure skidding, braking, and steering performance, and tire-road friction, engineers used those measurements to construct mathematical models of the cars' skidding behavior. One might consider this 'analysis,' but it was analysis in service of designing new systems that would mitigate the skidding (i.e., the wheel locking up) phenomena. This is just one way in which design processes produced new knowledge, and here that knowledge looked very much like scientific knowledge. Those models were then used to produce device designs that could modify appropriate model parameters in order to reduce the likelihood of any wheel locking up, and thus would reduce the extent of skidding. The case of antiskid design is an interesting one in part because the instruments that were designed to measure the phenomena because essential parts of the systems that were designed to mitigate skidding. Sensors and computers that measured and calculated angular velocity of wheels and could compare angular velocity to the vehicle's linear velocity and deceleration were critical technologies of both the design process and the design of the actual commercial systems. This overlap isn't a property of all designs, but it may be fairly common on certain kinds of devices, especially on technologies that move.

Engineers from several different companies and some public institutions modeled skidding mathematically and determined the parameters that affected it. Then they could then use that knowledge to design a system that would react when skidding was about to occur. The model of the phenomenon of skidding dovetailed right into the device designed to first detect imminent skidding, which could then send an electrical signal to the brakes to pulse them and allow a wheel to speed up and avoid locking up. Designing the devices really depended on determining

how the system would compare the data sent by the sensors. How much computational technology would the system need? There were largely mechanical designs, produced for example by Lucas Ltd., ranging to designs that used considerable numbers of electronic components to compare velocities, produced by Robert Bosch GmbH. Given the complexity of the underlying mathematical models, the systems really worked in different ways and were truly competing designs. Given the importance of the engineers from Bosch in the knowledge community, their designs had particular influence in the community and were widely accepted to be the state-of-the-art. Other engineers followed suit and designs incorporating electronic components became more and more common, to the point where being electronic was eventually taken to be part of the solution to the problem of designing antiskid systems. The electronic aspect of the design had to be sold to car manufacturers, who were skeptical of the cost and reliability of electronics in the early to mid-1970s (Johnson 2001). When many engineers in the knowledge community turned to favor electronic designs, the community opened up to electrical engineers who had knowledge that was now highly relevant to future designs, and revised electronics functions.

In the late 1960s and early 1970s several designs were produced that could be purchased on cars in the United States and western Europe. Designs in users' hands produced further feedback to engineers about how the designs worked in practice. Car dealers and repair shops were a conduit of information about the systems, their reliability, and problems users had. Sales were also an indicator of user interest in the system. Car manufacturers had a rule of thumb that buyers would be willing to pay up to about 10% additional, relative to the car's total cost, to buy a safer car that was less likely to skid. Users were assumed to be uninterested in systems that would, for example, double the cost of a car. Therefore, economic factors became a consideration in the design—but the different costs of different types of cars left a wide range for the cost of antiskid systems. Cost favored larger, more expensive cars as the first models to be equipped with antiskid devices and only with mass production would prices become manageable for less-expensive models. As a result, this also added the question of how to design systems that could eventually be manufactured more cheaply, especially where the electronics could be mass produced affordably. As shown by the story of the Model T, designing an artifact itself can be a wholly different project than designing how to mass produce it (Hounshell 1984). Most mass-produced technologies have this feature. Production is part of the design of the prototype; that is, production methods are considered. But once a prototype is produced and a decision is made to scale up production, then a second design process begins that considers how to make large numbers of the devices at a particular price. This is no less a design than the design of the initial prototype.

Designs and Assumptions About Users

What the case of antiskid system design reveals is the socially complicated process of defining a problem amenable to a technological solution, then figuring out what that problem really consisted of. Once a problem, skidding in this case, was narrowed down, then engineers working for different firms could design a technology to produce the desired phenomenon—here that was a car that avoided skidding under conditions that previously would have led to skidding, such as hard braking, ice, or wet pavement. This process of translating a function into a material design would look similar were the device a washing machine or a vacuum; it might look different for a computer since the function of a computer and the ways the consumer uses a computer are less tightly coupled together. The computer poses interesting questions for design because of the complicated and variable ways its function and its use are only loosely coupled. As a result, the nature of the function and the nature of the user matter profoundly to how designs are produced. Furthermore, the novelty of a technology also bears on the way its design

is conceived. Technologies that are not replacing or augmenting an existing device already have far more to explain to consumers; there's more speculation about how users will engage a device. Speculative designs (e.g., the Apple II, early smartphones, or even the early automobile) require engineers to make far more assumptions about the potential users and consumers of a technology; sometimes those assumptions are wrong and the design is modified then based on user feedback. The Model T offers a number of instances of user feedback that fine-tuned its design over twenty years. We think of the Model T of being one model—the same—over this period, yet many minor changes were made and the entire production design changed with the introduction of the assembly line in December, 1913 (Sullivan 2016). Designs are more than the outward appearance or some general account of how a device works; design also includes how a technology has to be made and how its users engage it.

Good Versus Bad Design

Although we think of design in aesthetic terms, good designs are often invisible. They're the designs that 'feel natural,' or 'seem intuitive,' those that cohere to prior experiences we've had so nothing seems to change, or if it does to just subtly improve. They don't cause accidents except for human error, and in fact ideally they even minimize human error. We appreciate good design of processes as well: e.g., production processes that minimize and control waste, software that allows students to sign up for classes effectively, division of motor vehicle protocols that minimize our time waiting, and interfaces that allow us to know that our taxes have submitted. Software designs of interfaces in particular have an aesthetic dimension as well as true functionality. Some are material, and some are virtual, but they're all processes where users—their biases, preferences, and sometimes their foibles—must be considered in the design. Designers must consider how users will misuse a system and ensure that the design allows such misuse without catastrophic consequences. In nearly all cases these complex technologies were designed by teams of people with different skills and knowledge sets and in the process of designing those professionals often revise what they know about the system and its users. That is, they produce and often share new knowledge about the processes. Understanding how this all happens is a serious philosophical question, with three axes: the material and aesthetic, the epistemic, and the social. When designs are successful, in many cases, we mark their success by rendering the design, and the designers, invisible.

References

- Baird, D. (2004). *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley, CA: University of California Press.
- Billington, D. (1983). *The Tower and the Bridge: The New Art of Structural Engineering*. Princeton, NJ: Princeton University Press.
- Bucciarelli, L. (1994). *Designing Engineers*. Cambridge, MA: MIT Press.
- Constant, E. (1987). The Social Locus of Technological Practice: Community, System, or Organization? In W. Bijker, T. Hughes, and T. Pinch (eds.) *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press.
- DeRidder, J. (2007). *Reconstructing Design, Explaining Artifacts*. Delft: Delft University Dissertation.
- Ferguson, E. (1994). *Engineering and the Minds' Eye*. Cambridge, MA: MIT Press.
- Hill, P. (1970). *The Science of Engineering Design*. New York: Holt, Rinehart and Winston, Inc.
- Hounshell, D. (1984). *From the American System to Mass Production, 1800–1932*. Baltimore, MD: Johns Hopkins University Press.
- Hughes, T. (1987). The Evolution of Large Technological Systems. In W. Bijker, T. Hughes, and T. Pinch (eds.) *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press.

- Johnson, A. (2001). Unpacking Reliability: Robert Bosch GmbH and the Construction of ABS as a Reliable Product. *History and Technology*, 17, 249–270.
- Johnson, A. (2009). *Hitting the Brakes: Engineering Design and the Production of Knowledge*. Durham, NC: Duke University Press.
- Kranakis, E. (1996). *Constructing a Bridge: An Exploration of Engineering Culture, Design, and Research in Nineteenth-Century France and America*. Cambridge, MA: MIT Press.
- Kroes, P. (2000). Engineering Design and the Empirical Turn in the Philosophy of Technology. In P. Kroes and A. Meijers (ed.) *The Empirical Turn in the Philosophy of Technology*. Amsterdam: Elsevier.
- Petroski, H. (1992). *To Engineer Is Human: The Role of Failure in Successful Design*. New York: Vintage.
- Petroski, H. (1999). *Invention by Design: How Engineers Move From Thought to Thing*. Cambridge, MA: Harvard University Press.
- Rosenberg, N., and Vincenti, W. (1978). *The Britannia Bridge: The Generation and Diffusion of Technological Knowledge*. Cambridge, MA: MIT Press.
- Sullivan, L. (1896). The Tall Office Building Artistically Considered. *Lippincott's*, 403–409.

Further Readings

- Bijker, W. (1997). *Of Bicycles, Bakelite, and Bulbs: Toward a Theory of Socio-Technical Change*. Cambridge, MA: MIT Press. A sociological account of the ways technologies' uses and their designs are interrelated.
- Koen, B. V. (2003). *Discussion of the Method: Conducting the Engineers' Approach to Problem Solving*. Oxford: Oxford University Press. A philosophy of the role of heuristic thinking in engineering design.
- Mitcham, C. (1994). *Thinking Through Technology: The Path Between Engineering and Philosophy*. Chicago: University of Chicago Press. A wide-ranging account of the relationship between engineering design and engineering science in the making of technology.
- Norman, D. (1988). *The Design of Everyday Things*. New York: Basic Books. A psychological account of the ways objects affect users and their capabilities.
- Vincenti, W. (1993). *What Engineers Know and How They Know It: Analytical Stories From Aeronautical History*. Baltimore, MD: Johns Hopkins University Press. A series of case studies about the way knowledge is produced in service of engineering design in aeronautics.
- Vinck, D. (2009). *Everyday Engineering: An Ethnography of Design and Innovation*. Cambridge, MA: MIT Press. A series of ethnographic studies about design engineers in real workplaces.

Dancing the Device

A Translational Approach to Technology

Albrecht Fritzsche

What if the engineer's perspective and the perspective of users of technology in nonindustrial application were radically different? What if they were related to incomparable cultural settings that do not allow a direct mapping from one to the other? Applications of technology outside of industry, as they are increasingly discussed due to the ongoing digitization, then require an additional, creative effort from the users. Translation studies allow us to interpret this effort as a form of paraphrasing technical operation in a new contextual setting. In addition, if the assumption of a radical difference is taken seriously, users of technology actually add something new to technology. Following Walter Benjamin's approach to translation, we can say that users open up paths for relations across boundaries without eliminating foreignness. Personal experience with technology can accordingly not be treated as a secondary phenomenon in the aftermath of technical operation. It has to be appreciated on the same level as the engineer's approach in the design and analysis of technology and its reflections in other disciplines. Metaphorically, one might say that engineers provide the music and users make the steps. Both have to be considered at the same time in order to gain an appropriate understanding of technology

Introduction

Digitization has made a lot of progress during the last years, but it has really not progressed in the way that one would have expected. Early examples in logistics and manufacturing suggested that industrial applications were the main driving forces; consumer applications with little connection to any professional activity received comparably little attention. This seemed to be a plausible approach. For a long time, the development of technology was driven forward by engineers in scientific or industrial research institutions that focused on highly structured application scenarios like manufacturing, traffic, warfare, or business communication. Steam engines were first introduced to factories, cars relied on existing road systems, and telephones were installed in public places and companies long before they reached private households. Information technology progressed in a similar way. Computers, cell phones, and digital communication networks were designed for professional work in companies, public offices, and research institutions. The development of consumer applications followed later. During recent years, however, the situation has significantly changed.

The changing dynamics of technology development can probably best be illustrated by a look at the work conditions in industry. Employees in a company are nowadays likely to have the latest smartphones, tablet computers, navigation devices, intelligent bracelets, and other wearables at home rather than at the office, and they hardly need experts any more to tell them how they are used. Twenty years ago, this would have very rarely been the case. Most employees would have first become acquainted with new technology in a work environment and would have gone through some kind of training program there. Government offices and industrial

associations regulated and controlled the introduction of new devices in- and outside industry. Today, a lot of the initiative has moved to private households and open communities on the internet where trained experts often play only a peripheral role, and they seem in many respects more interested in digitizing their personal lives than their professional ones. This is also the impression that one can get from the huge initiatives to support digitization processes in professional application environments that have been started by many countries, including Advanced Manufacturing Partnership in the United States, Germany's Industry 4.0, the British Catapult program, or Made in China 2025. All these initiatives have different agendas that consider the state of the industrial sector in the country. The general purpose, however, seems to be quite the same: foster a structural change that brings the latest technology back to industry.

First experiences with these programs show that this is easier said than done (Oks and Fritzsche 2016). Although many intelligent and responsive mobile devices have already been introduced, their impact on the surrounding structures seems fairly small in comparison to what happens outside industry. This can be explained by various constraining factors, including resilient traditional role models, institutional boundaries, intellectual property rights and legal restrictions, mission strategy, and a lack of motivation for change. However, there might also be a more fundamental difference to consider: there is no reason why digitization should necessarily have the same impact within industry that it has in other application environments. In fact, one might even ask whether there is a common background on which both scenarios can be satisfactorily discussed.

The following considerations are based on the hypothesis that there is indeed a radical difference between application environments for technology inside and outside of industry. Inasmuch as the design and construction of new technical devices is dominated by engineering as professional activity within an industrial context, one must therefore conclude that users outside industry have to invest special effort to come to terms with technology in their application environments, which seems in many ways comparable to a translation between cultures. Translation studies can therefore be expected to provide an interesting background for our inquiries.

Industry and Users

Theories in the field of economics connect technology with the transformation of goods in a production process. This process is described by a production function, which defines the relationship between the maximal technically feasible output and the minimal inputs needed to produce that output (Shephard 1970). Technical feasibility means that the production process is considered to be free of any further disturbances. It assumes that all operative problems that reduce efficiency have already been addressed and solved. What remains is the question of allocative efficiency, which concerns the question how the bundles of inputs should be arranged to maximize the profitability of the production process. Technology is, in this sense, inherently connected to industry. It is not a matter of private life, and both those who build technology and those who use it can be considered as professionals, acting on rational principles of efficiency and effectiveness.

The notion of technical feasibility in a production function creates a dichotomy between idealized formal representations of technical operations and their actual occurrences in the world. This has been quite controversially discussed from various perspectives (e.g., Samuelson 1966; Koopmans 1979; Gehrke and Lager 2000). It corresponds with the differentiation between the abstract concepts of determinate operation expressed in the notion of tools and machines and the behavior of material objects in practical application. Among others, von Foerster has addressed this idea in his notion of a trivial machine (von Foerster 1984). Hegel describes the consolidation of such abstractions with actual experience as a dialectic process.

This process, however, does not have to be set into motion every time that a tool or machine is applied, if the application is embedded in an environment that already externalizes indeterminacy (Fritzsche 2010).

Industrial structures provide good examples for such environments. In fact, the establishment of conditions under which work is performable in sequences of determinate operations can be considered one of the essential characteristics of industry. In this sense, industrialization can be described as the attempt to establish spaces in which the only thing that matters is technical operation and everything else can be ignored. In contrast to the idea that the main function of a system is the establishment of boundaries in order to form units of specialized activity (e.g., Luhmann 1995), the main function of industrial structure would accordingly be the establishment of a mapping of a real-world environment to a consistent formal representation.

If so, it does not make sense to draw lines between economy and technology or management and engineering. Rather, it is necessary to make a distinction between the expression of structure in a formal code and the implementation of a reference to it in the real world. Both activities involve managers just as much as engineers; and philosophers of technology have frequently emphasized the importance of collaboration across disciplinary boundaries (Michelfelder 2010). What remains unclear is whether this will bring us into closer contact with, as Carl Mitcham calls it, the “real world of technology,” or just “how that world is manifested in technological discourse” (Mitcham 1994: 135).

Such manifestations are not only found in factory production and logistics, but in many cases, they are also taken beyond the boundaries of conventional industrial structure. Interestingly enough, economics remains rather vague about this. Industrial production is assumed to be connected with consumption in markets, where consumers are expected to express consistent preferences in the choices that they make, without further details about how these preferences come into being. From a more practical point of view, business and management research strongly advocates a closer study of consumer behavior in order to understand the formation of these preferences and take them into consideration during production (e.g., Heinonen et al. 2010; Vargo and Lusch 2004). With increasing information about consumers, companies become able to find new access points in their lives where they can be supported with highly customized products. The business processes of the companies are expanded to include consumer activity; consumers turn into co-producers (Ranjan and Read 2014). In many respects, this can be interpreted as an attempt to circumvent markets. Consumers are integrated into industrial environments, and there seems to be no more need for a mechanism to adapt demand and supply based on the actual exchange of goods.

In reality, however, it seems that markets shift elsewhere and take on a different form. This also happens when public life is reorganized in a way that makes it more compatible with industrial structure—for example, in road traffic with its immense infrastructure and legal regulations. Vehicle designs, road networks, fuel supply, and many other aspects of traffic are highly interdependent and develop together as a whole. Drivers have to follow clear rules that constrain the way how they can use their vehicles, where they can use them, and even the range of purposes for which they may be used. Similar to an internal industrial structure, such an application environment allows experts to focus on abstract design questions such as automated driving, spending years to come up with feasible solutions. At the same time, however, technology evolves as well outside of these environments and creates a completely different dynamic. To name but a few examples, the internet allows the creation of community-based solutions for passenger transportation that question conventional ideas of individual traffic, and external mobile devices are connected with the internal technical architecture of cars to alter their behavior and expand their range of functions beyond areas that remain under regulative control.

Such market shifts can happen much more quickly than the changes induced by industrial and industry-like activities. Market shifts are more versatile and variable over time than any regulative mechanism, and they constantly open up new directions for innovation. Therefore, market shifts and not regulative measures are the phenomena that need to be researched in order to gain a better picture of the dynamics of digitization outside of industry. Translation as a cultural concept allows us to reflect the idea of an outside from an interesting new perspective.

Translation

Considering the huge importance that translation must have had for cultures that gained their wealth through trade, it received very little attention in early philosophical writings. One possible explanation is that it was considered as an operative problem of using a language that occurs after the expression of an idea. In this sense translation appears as a rhapsodic art that deserves less appreciation than original thought, as expressed in Plato's dialogue *Ion*. In a famous quote from "De Optimo Genere Oratorum," Cicero distinguishes two different approaches to translation, connected to the role of an interpreter in the exchange with people from foreign cultures on the one hand and the role of an orator who speaks to an audience on the other hand. According to Cicero, an orator is more advised to paraphrase the original wording than an interpreter. Cicero decides to take the approach of the orator, but there is no reason to assume that he did not appreciate the work of the interpreter as well. Interpreters played a huge role in the Roman Empire, and there is evidence that they were even used in meetings where the present parties had knowledge of the other one's language in order to establish distance and highlight one's own position (Kurz 1986).

The liberty to paraphrase has remained a major topic in reflections about the practice of translation (Bassnett 1980). Translations of poetry and other writings with a higher literary claim were in particular considered to be in need of paraphrasing in order to preserve the sense of the original. Romanticism turned attention back to the idea of establishing distance in translation. In particular, Schleiermacher was highly influential with his essay on translation, in which he advocated an approach that made clear to the readers that they were confronted a foreign text (Schleiermacher 2004). Fidelity to the original required not to bring the author to the reader, but the reader to the author; even if it made the text more difficult to understand, the readers were expected to benefit from experiencing the text as something that belonged from a foreign culture (Roessner and Italiano 2012).

This approach to translation is frequently related to the increased sensitivity for national identities in Europe after the Napoleonic Wars and its expression in language, which played a particularly important role in Germany, because the country was still divided into many independent regional states. Romanticism, however, was also inspired by another experience of foreignness, caused by the ongoing industrial revolution, which increasingly unfolded its own dynamic of growth, outside of traditional social and cultural structures. This offers the possibility to interpret Schleiermacher's text as a reference to a more fundamental dialectic movement, as described by Hegel in his *Phenomenology* a few years earlier. Anonymous factory production also shaped the experience of technology as something foreign and strange, something that followed its own rules without being constrained by the laws and customs of the people who applied the products. In this foreign culture, the engineer assumed the role of the author in creating meaningful expressions. They remained strange to audience and gave them a better understanding of their own identity in experiences of nature and emotion.

From a translational point of view, technology can accordingly be said to have reached its application domains in metaphrases of its industrial origin that continued to evoke feelings of strangeness and alienation far beyond the Romantic Movement (most famously in Heidegger

1993: 320). Various technical professions give account of the specific effort necessary to make sense of metaphrasal expressions of technology in the application domains. Chauffeurs, typists, and other skilled experts mediated between the usage of tools or machines and their application for a purpose. At the same time, handbooks, training courses, and other educational measures increased technical literacy in the general population to facilitate the adoption of new devices.

Today, we are confronted with the opposite phenomenon. Most tools and machines seem to fit seamlessly into our daily practices. Industrial output is customized; it is intentionally designed to be useful and easy to use; and the need for intermediaries and systematic training has significantly decreased. Modern information technology plays an important role in making this possible by adding different layers of symbolic representation to technical devices. Application environments are extended by a digital component that is not subjected to the limitations of material boundaries. The functional architecture that determines the physical construction of a device can be rearranged to form different operational pattern with a better fit to the concepts of instrumental action that users have already internalized.

Users therefore see buttons on a tablet screen that appear to move backwards when they are touched; they feel resistance of a lever when they push the pedals in a car and hear a voluminous engine sound when they accelerate. In reality, the tablet screen is two-dimensional; the pedal works electronically, and the resistance is generated pneumatically as a simulation of a mechanical process, just as the sound of the engine is often very carefully fashioned. This does not make the experience of using technical devices less real. Nothing changes in the relation between the user, the instrument, and the application environment. What really changes is the relation between the engineering process and the experience of the device in practice. Speaking in translational terms, engineering is paraphrased; it is translated into something that has largely the same sense, but deviates strongly from its original expression.

The fact that digitized technology has evolved more quickly outside of industrial structures than inside now starts to make sense. The closer technology remains to the production process, the lesser is the need for paraphrases. If the facilitation of paraphrases is one of the major functions of digital data processing, there is much more need for it in application domains than in industry.

Authorship

In the early twentieth century, Walter Benjamin proposed a change of perspective in the discourse on translation, leading away from the actual treatment of texts towards the role of the translator as an intermediary between cultures (Benjamin 1923). Translation is in this respect not secondary to the work of an author. It stands on its own as a different kind of creative achievement than the production of a text. Benjamin's essay has lately become very influential in the field of cultural studies, in particular though the works of Homi Bhabha and his concept of a third space between cultures that makes transformation and transgression possible (Bhabha 2004; Rutherford and Angela 1990). This concept breaks with the idea of a melting pot in which cultures that live together merge into one common unity. Instead, they are considered to preserve their individuality in the usage of translations to relate to each other, not only in terms of spoken or written language, but all forms of cultural expression.

Understood as mappings of an expression in one language to another expression in a different language with the same sense, translations are never fully satisfactory. What makes cultures different is the very fact that such attempts must ultimately fail. Translations therefore have a more fundamental function: to create paths on which different cultures can reach a common ground by suggesting ways how they can relate to each other despite all difference. Translations thus allow a coordination of thought and action across cultural boundaries that enable people to engage in common endeavors for everyone's benefit. The differences between the cultures

are not resolved. The foreignness remains, but translation makes it possible to appreciate one another in being foreign. In this, translations constitute truly creative acts.

Paraphrases of technology in application consequently need further attention. According to the previous considerations, they do not simply rearrange technical function according to different external constraints: they create references to technology that allow users to come to terms with it, although it remains a foreign entity to them as human beings. Such translations of technology are just as inadequate as conceptualizations of users in industry. They cannot give a satisfying account of the original object of reference, but they make it possible to appreciate it in a foreign environment.

Organ projection or extension is one of the most popular motifs in the philosophy of technology (Kapp 1877; Gehlen and Berger 1980). It is usually connected with the idea of forming a greater human being or compensating for human deficiencies with tools and machines. This makes sense only on the basis of an industrialized conception of the human body that makes it accessible as a technical system, expandable with further devices. It is also possible, however, to take the opposite approach by treating technical devices as parts of one's own corporal identity. By doing so, they become unique and irreplaceable, just like anything else that is part of one's personality. A tool or machine that is produced in thousands gets an identity, not because it is customized for the needs of one person, but also because the person makes it their own. For this person, it is irreplaceable. Any other part that may be used instead will never be the same. From an industrial point of view, it does not make any sense to treat the output of mass production this way. At the same time, however, there is no way to relate to mass production in terms of personal experience, either. This is achieved only by forming a personal attachment.

There are numerous other examples of such translations of technology. They find expressions in anthropomorphized robots, tuned cars, loyalty to outdated devices, and many other phenomena surrounding technology that are not satisfactorily accessible from an engineering point of view, but nevertheless play a huge part in life with technology. Bringing such phenomena to industry does not make sense. Nevertheless, they have a huge influence on the development of technology. The treatment of smartphones, tablets, and wearables seems to fall into the same category. They come closer to the notion of organs than most other technology. They accompany their users throughout the day and they react to voice commands and touch. Users put endless effort into the installation of additional software and store data on them that make them truly individual, and they are involved in many very personal activities and emotional experiences. In order to gain a better understanding of digitization and provide guidance for its further development, future research will have to find ways to give an appropriate account of these phenomena.

In this context, it may be helpful to refer to economic models of markets where production and consumption are set apart from each other. Such scenarios, in which producers simply make offerings and consumers are free to decide among them, are the most interesting for research. Whenever there is already a shared notion of problems and solutions, or means and ends, between both sides such that consumers have only one alternative for a rational choice, there is little need left to look beyond the boundaries of industry. Consumers must be assumed to have the liberty to develop their own notions of problems and solutions, which may or may not take suggestions from other parties into account, but always includes an individual creative act. This is the necessary precondition for any dynamic of technology development outside of industry.

The Dance

Several years ago, a German research project used a cultural approach to translation for the study of dance (Klein 2009). In particular, it focused on Argentine Tango with its wide repertoire

of different expression that can be flexibly combined (Klein 2012). Argentine Tango has lately become very popular all around the world and many communities of dancers have developed their own specific forms of expression and dancing practices. This makes the dance a perfect subject matter for translation studies, but also for innovation research (Fritzsche 2015). In many respects, the metaphor of the dance also seems helpful to summarize the discussion of technology on the background of translation theories that has been presented on the previous pages.

Imagine the engineer who designs technology as an artist who composes a piece of music. Imagine technicians as those who play the music, and industry as the overall environment in which this takes place. So far, users of technology have frequently been treated like the audience in a concert hall that sits in the dark and listens to the work of genius that is performed on stage. In many respects, this seems to describe the dynamics of technology development during the last centuries. Today, however, the situation seems to be a different: The users of technology have left the darkness. They do not sit still any more but start to dance, and this makes clear that they play a much more active role than it seemed before. Nevertheless, they do not replace professional engineers and technicians. They just add their own, individual facets to technology in the movements that they make. Each of these movements is its own interpretation of the music that is played on stage; each is different and expresses a personal interpretation of the music. But each of the dancers can justly claim to perform the dance, and if it were not for the dances, the music on its own would not make the whole event a dance party, no matter how well it is played.

Digitized technology permeates private life, and in this infusion, it has created a new dynamic in the development of technology that does not fit to industrial structure. The metaphor of the dance might help to gain a better understanding of what is going on. The shift from professional to private application contexts has turned attention to the efforts that are necessary for every human being to come to terms with technology. These efforts take place in a whole different sphere than engineering, just like the movements of a dance in comparison to the music that is played. But they have to be addressed as well to gain a comprehensive understanding of technology in the age of digitization.

In the future, we will have to take more care of the personal and idiosyncratic aspects of technology, the bonds that people create with technical devices, the efforts they spend to come to terms with it, and the individuality that is expressed in it. With the digitization of technology and its shift to private life, it will become increasingly difficult to make general judgements and provide ethical rules for the treatment of technical devices, because they can have a different meaning in the life of every single person that has engaged in its usage, even if the pattern of usage is completely the same according to all data that have been collected in its course. The difference does not lie in the operation of the device, but in its translation to a personal experience.

References

- Bassnett, S. (1980). *Translation studies*. London and New York: Methuen.
- Benjamin, W. (1923). *Die Aufgabe des Übersetzers. Gesammelte Schriften IV*. Frankfurt: Suhrkamp, 9–21.
- Bhabha, H. K. (2004). *The Location of Culture*. London: Routledge.
- Fritzsche, A. (2010). *Engineering Determinacy: The Exclusiveness of Technology and the Presence of the Indeterminate*. Dordrecht: Springer.
- Fritzsche, A. (2015). Innovation Tango: the Economic Potential of Cultural Translation. *British Academy of Management Conference (BAM)*, Portsmouth.
- Gehlen, A., and Berger, P. L. (1980). *Man in the Age of Technology*. New York: Columbia University Press.

- Gehrke, C., and Lager, C. (2000). Sraffian Political Economy. In *Encyclopedia of Political Economy*. New York: Routledge.
- Heidegger, M. (1993). The Question Concerning Technology. In D. Krell (ed.) *Basic Writings*. New York: Harper Collins Publishers, 307–342.
- Heinonen, K., Strandvik, T., Mickelsson, K.-J., Edvardsson, B., Sundström, E., and Andersson, P. (2010). A Customer-Dominant Logic of Service. *Journal of Service Management*, 21(4), 531–548.
- Kapp, E. (1877). *Grundlinien einer Philosophie der Technik: Zur Entstehungsgeschichte der Kultur aus neuen Gesichtspunkten*. Braunschweig: Westermann.
- Klein, G. (2009). Bodies in Translation. In G. Klein (ed.) *Tango als kulturelle Übersetzung. Tango in Translation. Tanz zwischen Medien, Kulturen, Kunst und Politik*. Bielefeld: Transcript, 15–39.
- Klein, G. (2012). Toward a Theory of Cultural Translation in Dance. In S. Manning and L. Ruprecht (eds.) *New German Dance Studies*. Champaign, University of Illinois Press, 247–258.
- Koopmans, T. C. (1979). Economics Among the Sciences. *The American Economic Review*, 69, 1–13.
- Kurz, I. (1986). Dolmetschen im alten Rom. *Babel*, 32, 215–220.
- Luhmann, N. (1995). *Social Systems*. Stanford, CA: Stanford University Press.
- Michelfelder, D. P. (2010). The Philosophy of Technology When ‘Things Ain’t What They Used to Be’. *Techné: Research in Philosophy and Technology*, 14(1), 60–68.
- Mitcham, C. (1994). *Thinking Through Technology: The Path Between Engineering and Philosophy*. Chicago: University of Chicago Press.
- Oks, S., and Fritzsche, A. (2016). An Application Map for Industrial Cyber-Physical Systems. In Jeschke, Brecher, Song, and Rawat (eds.) *Industrial Internet of Things: Cybermanufacturing Systems*. Springer.
- Ranjan, K. R., and Read, S. (2014). Value Co-Creation: Concept and Measurement. *Journal of the Academy of Marketing Science*, 1–26. DOI:10.1007/s11747-014-0397-2
- Roessner, M., and Italiano, F. (2012). Translation: An Introduction. Translation. Narration, Media and the Staging of Differences. In F. Italiano and M. Rössner (eds.) *Translation. Narration, Media and the Staging of Differences*. Bielefeld: Transcript, 9–18.
- Rutherford, J., and Angela, F. (1990). *Identity, Community, Culture, Difference*. London: Lawrence and Wishart.
- Samuelson, P. A. (1966). Rejoinder: Agreements, Disagreements, Doubts, and the Case of Induced Harrod-Neutral Technical Change. *Review of Economics and Statistics*, 48(4), 444–448.
- Schleiermacher, F. (2004). On the Different Methods of Translating. In L. Venuti (ed.) *The Translation Studies Reader*, 2nd edition. New York: Routledge.
- Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton, NJ: Princeton University Press.
- Vargo, S. L., and Lusch, R. F. (2004). Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, 68(1), 1–17.
- von Foerster, H. (1984). Principles of Self-Organization—in a Socio-Managerial Context. In H. Ulrich and G. J. B. Probst (eds.) *Self-Organization and Management of Social Systems*. Berlin: Springer, 2–24.

On Harnessing of Birth in the Technical Age

Dana S. Belu

Technological intervention in human reproduction is a new norm. In this chapter I focus on reproductive technophilia and reproductive technophobia in childbirth. I interpret technological and nontechnological, but techné centered, childbirth through a phenomenological lens. Nontechnological childbirth is not synonymous with natural childbirth insofar as the latter (usually) refers to an automatic, biological process that simply overtakes the woman and activates a putatively innate birthing knowledge. I eschew this essentialist and patriarchal interpretation because it collapses childbirth into a passive, bodily function rather than viewing it as active work (Held 1989: 364). My working assumption is that all approaches to childbirth are always already culturally conditioned and involve some deliberation and choice—making by the pregnant woman.

Despite important differences, technological and nontechnological childbirth reflect contemporary Western, cultural norms of order(ability), control, and optimization. The technological model is technocratic because it

defines the body as a machine . . . and insists that, like a machine in a factory, the laboring body should produce its product within a specified amount of time; and if not, this birthing machine is obviously dysfunctional and in need of . . . repair.

(Davis-Floyd and Dumit 1998: 259)

Its antitechnological counterpart, Lamaze, traditionally emphasizes acquiring birthing skills and minimizing technical intervention(s). However, like its technical counterpart, it strives to conquer nature.

While both models reflect traditional masculinist commitments to aggression and control, the focus of my chapter is to rethink these childbirth scripts with the help of Heidegger's theory of technology. Applying Heidegger's concept of enframing (or *Gestell*) I advance a phenomenological interpretation that shows how these two childbirth models help to disclose the technological age and do not preexist it.

Heidegger's Theory of Technology: The World as Resource

In "The Question Concerning Technology" (1977 [1954]) Heidegger uses the term *enframing* to capture what he considers to be the essence of the technical age. His ontological, rather than sociological, account claims that enframing is "nothing technological" but an historical¹ (*geschichtlich*) mode of revealing (*aletheuein*), unique to the West. It refers to an inherited cultural outlook dominant in advanced, industrialized countries, according to which nature and people are reduced to raw materials, fungible units of consumption and commerce. Enframing naturalizes a value system geared towards continuous production, efficiency, and enhancement.

Heidegger says that in the age of enframing, people's lives increasingly lack rootedness and historical belonging or maternal protection (*Hut*) (Heidegger 1949: 46).

According to this theory, things show up only insofar as they have the potential to be ordered, i.e., "regulated and secured" (Heidegger 1977 [1954]: 175). The phenomenology of enframing² describes a relationship between an attitude of imposition, or "challenging-forth," and what this attitude discloses—the world as a heap of fungible raw materials, resources or "standing-reserve" (*Bestand*) awaiting optimization. The dominant value embodied by the attitude of challenging-forth is a constant "driving on to the maximum yield at the minimum expense" (Heidegger 1977 [1954]: 15). Heidegger says: "Everywhere, everything is ordered to stand by, to be immediately at hand, indeed to stand there just so that it may be on call for a further ordering" (Heidegger 1977 [1954]: 15). Enframing "orders what presences as available and reportable . . . as standing-reserve" (Heidegger 1949: 40). People and nature are reduced, each in their own way, to units of energy and information that are then "unlocked, transformed, stored, distributed and switched about" (Heidegger 1977 [1954]: 16). What doesn't show up as a resource or a potential resource (i.e., what cannot be challenged-forth) simply does not show up at all.

According to Heidegger's famous phenomenological analysis of the power plant on the Rhine (Heidegger 1977: 16), the river no longer gathers local traditions and lore, but is primarily seen as a water resource for the power plant. It is built into the dam, rather than the dam being built into the river. Although multiple technologies are used, challenging-forth is not, as we will shortly see, restricted to technological contexts.³

Heidegger's phenomenological approach peels back the layers of the mainstream, instrumental understanding of technology that sees technology as a neutral tool. This instrumental attitude is "limited to a subjective set of conditions" (Heidegger 1977 [1954]: 4–6) that it presupposes but cannot explain. One such (significant) subjective condition is the still-dominant, seventeenth-century Cartesian and mechanistic assumption that individual rational subjects control objects: i.e., those entities perceived to lack reason and agency. The mechanistic assumption that reason and "motion [were] external to matter"⁴ promoted a widespread instrumental attitude still popular today (Merchant 1989: 204).

However, since instrumentalism is stuck in modern subjectivism, it is unable to recognize this historical bias. For example, it cannot tell us why rationality, objectivity, and maximizing output matter but simply tells us that they matter. It cannot explain why today, for instance, norms such as efficiency, order(ability), and control are dominant and not others, such as tranquility, piety, or spontaneity.⁵ The instrumental subject fails to notice that what they do not master is the 'will to mastery' and that in multiple contexts the individual now treats him- or herself as an object. Iain Thomson puts it as follows:

Once modern subjects dominating an objective world begin treating themselves as objects, the subject/object distinction itself is undermined, and the subject is thereby put on the path toward becoming just another resource to be optimized, that is secured and ordered for the sake of flexible use.

(Thomson 2005: 60)

This result is mirrored in how entities are disclosed. Heidegger says:

Thus when man, investigating, observing, ensnares nature as an area of his own conceiving, he has already been claimed by a way of revealing that challenges him to approach nature as an object of research, until even the object disappears into the objectlessness of standing-reserve.

(Heidegger 1977 [1954]: 19)

Yet, another way is possible. Heidegger writes:

The field . . . appears differently than it did when to set in order still meant to take care of and to maintain [*hegen und pflegen*]. The work of the peasant does not challenge the soil of the field. In the sowing of the grain it places the seed in the keeping of the forces of growth and watches over its increases.

(Heidegger 1977 [1954]: 14)

This nonenframed approach points to a “technique that works with nature and the material, allowing beings to come forth at any given time, while modern technique strives to order, confine, control, and then to ‘challenge’ nature to produce” (Klawiter 1990: 71). In “Science and Reflection,” Heidegger clarifies this phenomenological difference in our understanding of nature when he says that nature as resource “is only one way in which nature exhibits itself” (Heidegger 1977 [1954]: 116). Human beings order and drive nature and technology forward but never produce how nature (or being) reveals itself. We always presuppose a specific understanding of nature, and we do this in our practices and not just in our heads. Blindness to this presupposition is a symptom of the modern subjectivist bias, a lack of awareness that we are historically determined to relate to nature in a controlling and mechanistic way.⁶ This unacknowledged limit of the instrumental position covers up and simultaneously points to the enframing.

According to Heidegger the instrumental understanding gets stuck in the technologies themselves and is therefore phenomenologically naive, providing only a “surface definition of technology” (Ihde 2010: 31). Two examples of such phenomenological naive views are technophilia, the unqualified love of technology, and technophobia, the fear and hatred of all things technological. Although seemingly different, these views are fundamentally similar because both see technology along instrumentalist lines, as merely a neutral object pushed around by a rational subject. Heidegger critiques these views as reactionary and avoids discussing the pleasures and terrors of technology they, respectively, endorse. He does, however, state that technophilia and technophobia perpetuate the “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” (Heidegger 1977 [1954]: 25). In the arena of reproduction, these views are embodied in the technocratic and nontechnological models of childbirth.

Technophilic Childbirth and the Call for Pain Relief

In the next two sections I situate Heidegger’s phenomenology, as described in the last section, in a feminist context. In my feminist phenomenology I show how the use of modern childbirth tools disclose women’s reproductive bodies, first as objects and later as resources.

It is undeniable that some advances in modern reproductive technology have made childbirth, for scores of women and newborns, much safer than it was in previous centuries. However, these advances have also promoted a social infatuation and overdependence on reproductive technologies, especially drugs, that often interfere with the well-being of mother and child. The increased use of advanced reproductive technologies (ARTs) continues to shift attention away from (the autonomy of) the mother and onto the newborn, whose safe arrival into the world is the primary medical objective. Realizing this objective requires the use of multiple technological interventions.

According to feminist Suzanne Arms, “The predominant technology in childbirth from the beginning of the 1850’s to today has been the use of drugs” (Arms 1994: 74). Today, the high rates of Cesarean births in the United States (Michaels 2014: 127), especially elective Cesareans

(the most enframed of all births⁷), mark the triumph of drug technologies, the marginalization of a pregnant woman's birthing techné or "know-how." My feminist phenomenological interpretation of the history of that loss notes the inversely proportional correlation between the medical (re)ordering of birth and women's marginalization from birth.

Between 1850 and the late 1960s new medical technologies in the United States and Europe changed birth from a woman-centered process to a technology-centered process that focused on producing tools, developing experimental techniques, and placing them in the hands of male obstetricians and their medical staff. Traditional childbirth with a midwife was on its way out. Along with it, woman-centered childbirth was also being marginalized; i.e., the experienced hands of (mostly female) midwives working with the laboring woman to bring a baby forth were displaced in favor of male-driven, theoretical medical knowledge. This knowledge enjoyed high social status but lacked experience, relied on poor theoretical instruction, and used devices to challenge-forth the birth (Louden 2002: 297). Moreover, the love of obstetric tools, coupled with the perception that most pregnant women are "unable" to birth naturally (i.e., without technical intervention), by the beginning of the twentieth century had framed birth in the United States as a pathological event. The cure for this pathology was believed to depend on using more medical technology. This view garnered support from, as it also reinforced, a growing technophilic attitude. However, after multiple medical interventions, childbirth become more technologized but no 'cure' was found. What was found, however, were multiple ways to enhance⁸ or optimize it: i.e., to switch it about, regulate it, secure it, and expedite it (Heidegger 1977 [1954]: 16) for purposes that often serve neither mother nor child and transform birth into an "obstetrical manufacturing process" (Chamberlain 1998: 169).

Modern childbirth is believed to have started in the 1730s when a Scottish apothecary named William Smellie perfected the obstetric forceps (Rich 1986: 141–142). For decades the forceps were used sparingly and successfully, often saving both the baby's life and that of the mother in emergency situations. By the end of the eighteenth century and beginning of the nineteenth century, however, the forceps were used for a variety of reasons, to ease the tedium of delivery, to save the surgeon's time, and simply because surgeons were infatuated with a new technology that symbolized progress. This overuse often resulted in botched births, damaged mothers and dead babies (Wajcman 1991: 65). Nonetheless, the technology was out there and could not be put back in the box. In this respect it was out of control.

In *Recreating Motherhood*, Barbara Katz-Rothman writes: "When childbirth became a medical event, women lost control over their own birth experiences. The medicalization began with the eradication of midwifery as a profession, and continued largely unabated until the home birth and midwifery movement of the 1970s" (Katz-Rothman 2000: 154). In *Of Woman Born*, Adrienne Rich underscores this view: "No one disputes that within recorded history, until the eighteenth century, childbirth was overwhelmingly the province of women" (Rich 1986: 131).

Although midwives controlled the birth process, this control did not challenge the hegemony of patriarchy, which exerted a stringent imposition of motherhood on all women and control over the fate of their offspring that continued unabated. Even the midwives' control over childbirth was not to last. By the middle of the nineteenth century, most women in the United States were no longer perceived as being "capable of standing the pain and stress of labor." Thus, if "women were no longer able to withstand the pain of labor, new methods . . . were needed" (Louden 2002: 343). By the beginning of the twentieth century, three interconnected features defined modern childbirth (and continue to do so today): "new methods of pain relief, massive intervention in normal labors, and admission to hospitals for as many cases as possible" (Louden 2002: 343).

The beginning of a biomedically constructed birth can be traced back to the marginalization of the midwife in Europe and the simultaneous rise of forceps overuse. This overuse was

initially (and remained primarily) connected to the sheer love and control of technological innovation by male doctors, and later to the doctor's desire to "expedite" (Heidegger 1977 [1954]: 15) birth. According to multiple sources, including the disparate works of Loudon, Rich, and Wajcman, this overuse was not primarily motivated by a concern with relieving women's pain in labor. A lack of concern with women's well-being, especially poor and working-class women, is not at all surprising considering the deeply patriarchal and classist norms at the time. In fact, monopoly of the forceps was tightly guarded during the seventeenth century by its male inventors—the Chamberlain father and son team—who used it selectively, and for a high price. Midwives, who were almost exclusively female, were prohibited from using technology during childbirth, including the forceps.

According to my Heideggerian interpretation, the overuse of the forceps begins to frame the birthing woman as "an object of research" (Heidegger 1977 [1954]: 19) who primarily served the technology rather than the technology serving her interests. Nor did it seem to serve the interests of the obstetrician *qua* obstetrician because he was (mostly) not alleviating pain but causing it, often along with severe injury⁹ to both mother and child. Moreover, the high failure rate of the technology did not deter its use. Its use remained controversial through the nineteenth and twentieth centuries.¹⁰ Today, the use of forceps has sharply declined, being widely replaced by vacuum extractions.

However, in the nineteenth century, the forceps was popular and often coupled with the use of ether and chloroform. By the early nineteenth century, two male physicians, Clement and Leake, definitively "establish[ed] the lithotomy (lying-down, therefore passive) position as the preferred one for women in labor" (Rich 1986: 146). Then, in 1847, Sir James Young Simpson discovered the anesthetic properties of chloroform (Arms 1994: 55). This technology consolidated the (male) physician's control of birth by rendering the woman useless and unconscious, a mere medical resource. The drug erased her pain and memory of birth. Adrienne Rich describes the experience in the following words: "Floating into euphoria as she reclined on her back on the doctor's table, a woman yielded everything—her attention, her effort, her responsibility, and her sense of protection toward her baby—to the authority of the physician" (Rich 1986: 56). Rich sums up the patriarchal demeanor of this procedure as follows:

At the onset of labor, the woman was placed in the lithotomic (supine) position, chloroformed, and turned into the completely passive body on which the obstetrician could perform as on a mannequin. The labor room became an operating theater, and childbirth a medical drama with the physician as its hero.

(Rich 1986: 170)

But more is going on. Rich's eloquent account assumes a Cartesian, subject/object framework wherein the physician is cast in the role of a rational yet unfeeling doctor and the woman is relegated to the object position. However, in view of the Heideggerian enframing, the real hero in Rich's quote appears to be not the physician but the drug itself, whose deployment reframes both physician and woman as resources. If we view chloroform on a spectrum with its significant successors—twilight sleep and the epidural—we notice how the use of chloroform already discloses the woman as more fungible than did the limited use of the forceps, something her body could oppose or reject.

Using Heidegger's account of enframing (as described in the previous section) allows us to see the difference between the body as object and the body as resource. Remember that according to Heidegger, one of the essential features of the technical age is that entities are no longer, as in modernity, disclosed as objects or things that stand over and against subjects (Heidegger 1977[1954]: 17), defined by relatively fixed boundaries, (social) uses and purposes. A resource

is an object that “disappear[s] into the objectlessness of standing-reserve” (Heidegger 1977 [1954]: 18) and is pliant, expeditious (i.e., it “drives on to the maximum yield at the minimum expense”), open to multiple purposes, and serves a (technical) network that drives toward continuous enhancement. The use of chloroform reveals many of these features.

For instance, chloroform enhances the obstetrician’s access to women’s reproductive bodies in a qualitatively deeper way than did the forceps before it. While it removed the pain of childbirth, thereby satisfying the desires of many women at the time, it also rendered them unconscious, thereby opening more medically experimental opportunities for the physicians. The withdrawal of the woman’s moving, affective body enabled physicians to experiment with the Cesarean technique. In fact, chloroform contributed to the expansion of modern Cesarean surgery,¹¹ which, in turn, led to a mass preference for physician-controlled birth in a hospital setting. However, one significant limit remained to be reckoned with: the body’s high sensitivity to death from overdosing. Due to high numbers of accidental mother and infant deaths, chloroform (in birth) paved the way for twilight sleep before it was finally discontinued.

The use of twilight sleep framed the pregnant woman as object. By 1914 and before terrible side effects surfaced, twilight sleep, composed of morphine and scopolamine, became American women’s preferred drug for both Cesarean surgeries and nonsurgical hospital births. It offered the same benefits as chloroform and did not accidentally kill as many women. However, due to, in part, a lack of proper medical training in administration and monitoring, it induced hallucinations and violent outbursts in many women. Women’s wrists, elbows, and knees were cuffed to their hospital beds to prevent self-injury and injury to the fetus and staff (Arms 1994: 78). In order to conceal the cuff scars from the husbands, who were not allowed in the hospital rooms, the hospital staff switched to lambskin straps. Even with their use some women wore away much of the skin around their wrists from thrashing (Michaels 2014: 15). Many women were routinely left alone for hours, tied to their hospital beds, sometimes in soiled sheets.

During a nonsurgical, twilight birth the nurse is routinely asked to press, even lean, on the mother’s belly in an effort to push the baby out while the physician pulls the baby out of the birth canal with the forceps. Whether the baby was pulled out or delivered via Cesarean, twilight sleep, like chloroform before it, ensured that the woman was not a conscious partner in birthing her child(ren). According to Sunnye Strickland, a pioneer in natural childbirth, many obstetricians “had not seen women awake for a birth except as a mistake, either a precipitous delivery or their own late arrival at the hospital” (Zwelling 2001: 17–21). This was reported to be the case as late as the late 1960s. Women’s disembodied experience of labor had become the norm.

Two problematic aspects of twilight sleep were the mother’s birth amnesia and the drug’s adverse affects on the newborn’s central nervous system. Many women felt bewildered and upset when presented with a newborn whom they had no memory of birthing. Some women reported thinking that they were still pregnant. Many of the newborns were drowsy and some needed resuscitation. Despite the adverse side effects, twilight sleep was used in the United States until the early 1970s.

From this brief overview we can see that twilight sleep does not ultimately serve the woman or the fetus but rather both of them serve the drug and its relationship to a hospital network that orders physicians and nurses to expedite birth. The drug is supposed to serve multiple ends: i.e., removing pain, controlling the process, and rendering the body a docile resource for obstetric intervention. However, docility is not achieved. Too often, and unpredictably, the woman strikes back. Her physical and psychological reactions are messy and obtrusive, reinserting her agency and affect, however poorly, into the process. These responses show her as object but not as resource. As we will shortly see, the use of the epidural demotes the position of the

obstetrician as subject set over and against the woman as object, and succeeds where chloroform and twilight sleep failed. It renders the woman quiet, conscious, and docile.

In the mid 1960s, the electronic fetal monitor (EFM), and the first version of the spinal epidural were widely introduced in hospitals throughout the United States. The EFM band encircles the woman's belly and restricts her movement, while the EFM machine prints out a strip with markings that represent the baby's heartbeat. Despite the fact that "numerous studies have shown that continuous EFM monitoring does not improve birth outcome" (Cartwright 1998: 244, 250), pregnant women continue to believe that the EFM monitor somehow protects the fetus (Cartwright 1998: 248). The actual benefits of EFM are to protect doctors in court, promote hallway obstetric conversation, and allow for remote physician monitoring (Cartwright 1998: 244–245).

When coupled with EFM, the epidural foregrounds the resource status of the woman's reproductive body. The epidural is "an anesthetic injected into a space inside the spine, blocking nerve transmission. Epidurals were first touted by doctors and nurses because they made for quieter labor wards than did twilight sleep" (Arms 1994: 80). In addition to alleviating pain and inducing a state of euphoria, it allowed mothers to remain awake during surgery (Aken and Gogarten 2000: 772)¹² and to immediately hold their babies afterward. Epidurals reduced pain and the number and intensity of contractions. Intense contractions were usually brought on by the drug pitocin, administered to catalyze labor or to activate a stalled labor. The drug is administered through a heparin lock, a collection of small tubes attached to a catheter and inserted into the arm. The heparin lock is a preemptive measure, is mandatory in most hospital births, and is a central part of the technical network that promotes a biomedically constructed, technocratic birth.

Epidurals slow down labor and diminish feeling below the ribs, compromising a woman's ability to push her baby out without more technological interventions, such as episiotomies, vacuum extractions, or Cesarean surgery. The recent increase in the use of epidurals (Michaels 2014: 11, 144) seems to have contributed to an increased rate in Cesarean sections,¹³ especially elective Cesareans, the most efficient delivery option today.

The epidural reveals the woman's reproductive body as a resource. It renders this body fungible while itself withdrawing, remaining concealed. It can be administered continuously and in various amounts to suit individual preferences. It is flexible, efficient, and carries little to no risk of death. Because the woman remains awake it gives the impression of birthing 'normalcy,' by comfortably mimicking natural birth and thereby enhancing 'docility' or the body's capacity to be "subjected, used, transformed and improved" (Foucault 1977: 136). By comparison, chloroform and twilight sleep appear intrusive and primitive. Thus, the more invisible and 'normalizing' the technology the more fungible the body becomes. The body's resistance, as an object—the unruly, embarrassing 'corpse' of the still-living woman—has now disappeared.

The epidural also reveals the woman as a resource because, as I noted in the first section, it allows her to treat herself as object. She is aware that things are happening to her but she does not feel any pain, pleasure, or effort, all of which would animate the experience and give it personal traction. Many women feel "detached from the experience of birth" (Michaels 2014: 129) and this renders the experience anonymous. Viewed alone, the epidural goes far in making what I call *reproductive enframing* visible. Reproductive enframing¹⁴ aims to capture the normalization of the medical interrogation of women's reproductive bodies that casts them as fungible. However, reproductive enframing in childbirth is most visible as the ensemble of hospital drugs and technologies working together.

The technical 'interlocking paths' formed at a minimum, by the hospital clock, the heparin lock, pitocin, and the EFM presuppose the enforced technique of the lithotomic position and culminate in the use of epidural anesthesia. Together they transform childbirth into an

‘obstetrical manufacturing process’ that is primarily served by the woman’s reproductive body rather than serving it. Hospital birth is closely monitored. If ‘progress’ does not fit into a narrow time frame, usually twenty-four hours or less, physicians will pressure women to have a Cesarean. A constant reliance on technical intervention would seem to promote the deskilling of the obstetrician’s hands-on expertise necessary to help with the twists and turns of labor. This makes more technical intervention into an attractive option. In this picture women and their doctors become part of a medical network that neither one of them controls. While the doctor can still be seen to animate the technical network, they are in fact ordered around by webs of administrative protocol and insurance liabilities that also frame the doctor as a resource. The erstwhile subject-object relationship no longer holds.

Radical and socialist feminists¹⁵ point out that the reason behind women’s choice of medicalized birth is the internalization of patriarchal norms that have not been successfully identified and/or resisted.¹⁶ In other words, too many women fail to think critically about the patriarchal norms and values they identify with, thus reproducing a patriarchal synecdoche, i.e., reducing women to their reproductive bodies and exploiting these bodies. For instance, Barbara Katz-Rothman’s influential book *Recreating Motherhood* is predicated on exposing the twin domination of pregnancy and childbirth by patriarchal ideology and technological ideology. The former sees the child as a “product . . . a seed planted in woman” while according to the latter, “technology is used to problematically frame mother and fetus as separate beings; nesting Russian dolls, one inside the other” (Katz-Rothman 2000: 160).

However, patriarchy alone cannot tell us why the norms of order, efficiency, and control matter; it simply tells us that they matter. Male and (female) obstetricians today are ordered to implement technological intervention by anonymous hospital policies, a network of insurance liabilities, and customer satisfaction that no one controls but that control everything. Thus, a normative desire for order(ability) and control is deployed in the medical network as the network itself, which, as Heidegger says, makes sure that the “regulating itself is, for its part, everywhere secured” (Heidegger 1977 [1954]: 16).

Technophobic Childbirth and the Upside of Pain

Despite today’s increase in technological births¹⁷ recent history in the United States shows a resistance to this technological trend. Starting in the 1960s and through the late 1970s, American women pushed back with the help of the Lamaze method. Lamaze taught them techniques for birthing without technology. This pushback made progress for a while, especially during the rise of social movements such as feminism in the United States. However, in my Heideggerian reading, this method anticipates central features of the enframing.

I suggest that women today could benefit from thinking about drug-induced technological birth as being fundamentally similar to the drug-free or low-tech alternative. Neither one achieves, what Heidegger calls, “a free relationship to technology” (Heidegger 1977 [1954]: 4, 34). The traditional and technophobic Lamaze method reflects—similar to the technophilic method it criticizes—some of the norms of enframing, including orderability and control. These two approaches can be seen to express different stages of reproductive enframing. The over-medicalized approach enframes by inserting the pregnant woman into the system as an object of technical action. The Lamaze approach enframes by making the pregnant woman the subject of technical action, a kind of laborer who is working to give birth. However, since both approaches presuppose order, control, and optimization, they project, each in their own way, the respective subject and object as resources.

The popularity of Grantley Dick-Reed’s *Childbirth Without Fear* (1933) and Fernand Lamaze’s *Painless Childbirth* (1956) promoted a nonmedical and somewhat technophobic

approach to childbirth.¹⁸ I am not suggesting that all or even that most women in the United States subscribed to Lamaze during the 1960s and 1970s, but enough women did that it became a dominant and recognizable trend. However, it was not to last: Lamaze is still being practiced, but it has been pushed from the center to the margins.

With the Lamaze method, women learned psychological and physical skills for birthing without drugs and medical technologies. This restored the dignity and power they had lost during the previous decades. It wrested some of the control of birth from the obstetricians back into their own hands, albeit for a short time. Although most women used the Lamaze method in a hospital, in a birth center housed within a hospital, or in a freestanding birth center (Michaels 2014: 132)¹⁹ where medical technology was available if needed, few women actually used the technology. Instead they relied on their acquired skill, trusted their bodies, and worked with their partners to see the birth through.

Fernand Lamaze became acquainted with the psychoprophylactic method, influenced by Pavlovian conditioning, on his trip to Russia in 1951. There he learned that most women labored naturally in hospitals and without the use of drugs that were so prevalent and, in his opinion, harmful, in the West, especially in the United States. That is, he learned that labor was work (Lamaze 1970 [1958]: 25, 32), something a woman actively and consciously does herself—learning to breathe with a great deal of control and precision, avoiding passivity and relaxation as she increases sensitivity to pain. The goal is to recondition the impulses of the nervous system by changing deeply entrenched attitudes, language and behavior.²⁰ A medical techné is certainly used, but almost always without technologies. This techné gives the woman a heightened sense of control and mastery over her body. According to Lamaze:

The Read method is based on the theory of eliminating tensions caused by fear, and thereby letting nature take its course unhampered by harmful emotions. The Pavlov method, while it agrees with the principle of *conquering fear by knowledge*, also makes use of conscious mental and physical control of the birth process. This control is attained through exercises and education designed to build conditioned reflexes which will stand up during the stress of labor and enable the woman to direct her own delivery. . . . That is why we do not call our system natural childbirth. *The final result should be better than nature.*

(Lamaze 1970 [1958]: 27 my emphasis)

According to this subject-object, Cartesian model, an anonymous and elusive ‘nature’ is controlled and optimized through a personalized integration of technique rather than being externally regulated by technology.²¹ In a Soviet-styled, masculinist framework the woman shows up as a laborer, a self-possessed subject. Her mind and body are the instruments of her technique. She has become a subject who is confident in her ability to control her body, to stay on task and to “regulate and secure” her birthing progress from within. Lamaze describes this techné as one that is predicated on the development of “rational knowledge, analysis and control” (Lamaze 1970 [1958]: 110). As one doctor put it: “Lamaze’s emphasis on discipline and mastery are well suited for someone who likes to be in control. The ideal Lamaze woman is, in fact, like a superbly trained athlete who has disciplined herself to perform under intense pressure” (Verny 1981: 110).

The Lamaze woman is caught up in a paradigm of performance and expertise. She plans and trains in order to conquer her fear of pain and ultimately to conquer pain. Instead of simply accepting the suffering that may come with childbirth or seeking refuge in drugs, she now triumphs over the pain through discipline and self-control. The practice of Lamaze illustrates an early stage of the reproductive enframing. The woman’s work is organized by the desire to order and control the birth process as closely as possible and to avoid the hospital’s technocracy.

Instead of these norms being imposed upon her, she shows up as the subject who actively imposes them upon herself, her body. Unlike the self-imposition of the woman who opts for the epidural, however, this self-imposition is engaged, skilled, and empowering. The presence of her Lamaze coach and partner underscores her new position as subject in a hospital setting where she used to be a mere object pushed around by doctors.²²

However, a closer phenomenological scrutiny reveals this subject to be, in fact, lacking control in important ways. The Lamaze woman has limited control over ‘nature’ going wrong. If things go ‘wrong,’ her *techné* is not able to handle the medical complications and emergencies that may arise. If the technique does not work for her, she may have to endure the pain and, in the traditional Lamaze model, be dissuaded from asking for pain relief drugs by her coach (Michaels 2014: 125).

Lamaze (often) makes for a slow birth process. It is not efficient.²³ Technological devices compromise a woman’s *techné* but quicken the process with plenty of precision. Many women today seem to just want birth to be ‘over with’ so they can see the baby. It seems that compromised efficiency and control are chief contributing factors why low-tech and drug-free childbirth continues to wane (Michaels 2014: 143) while the number of techno-centric births skyrocket. In fact, in order to survive in a technocratic market, Lamaze has, in the last couple of decades, accommodated some technical intervention and has softened its standard critique of natural home birth as lacking in standards, skill, and commitment (Michaels 2014: 143–154).

Despite Lamaze’s promise of an empowered and painless birth, it continues to be marginalized by its technocratic counterpart and its external forms of control. I have suggested that the main reasons for this displacement have been inefficiency and a restriction of control. However, because (traditional) Lamaze strives for maximum control and optimization, without advocating slowness but merely accepting it as a possible side effect, it contributes to the disclosure of the technological world.

Concluding Reflections

In recent decades, feminists have criticized growing technological intervention in childbirth. Many allege that it is another form of patriarchal oppression of women’s bodies and agency for which the appropriate antidote is a wholesale shunning of reproductive technology and a return to natural, drug-free childbirth.²⁴ While I recognize the legitimacy of this concern, I suggest that both technological and low-tech or no-tech options, like Lamaze, presuppose and promote order(ability) and control.

With its emphasis on restoring women’s dignity and acquiring skills that will help the mind to control the pains of the body, (traditional) Lamaze reflects unquestioned Cartesian, subject-object assumptions. However, the Lamaze woman’s self-imposition of a training regimen that will order and optimize her body brings in view the collapse of the subject-object distinction and is characteristic of the Heideggerean enframing. This collapse is more visible in the typical (technophilic) hospital birth where women’s reproductive bodies are harnessed by a variety of technologies intended to produce an ordered and orderly birth. However, the low-tech/no-tech instrumental attitude of the Lamaze woman, directed at her body, already helps to, paradoxically, provide a glimpse into this collapse.

By combining Heidegger’s theory of technology with feminist phenomenology I tried to show that when choosing between technological and nontechnological (or low-tech) birth we may be choosing the same: order(ability), control, and optimization. Thus, one option (too often) masquerades as two deceptively different choices. However, once we are able to understand how the reproductive enframing influences our birthing choices then we may come one step closer to developing a free relationship to reproductive technologies.

Notes

- 1 This sense of history is ontological and it refers to what Heidegger calls “the history of Being” rather than to a sociological and empirical account of events produced by human beings. Heidegger calls the latter “historiography.” In the ontological sense, history constitutes discrete horizons of meaning, first articulated in Plato’s metaphysics. We inherit these horizons of meaning and they serve as the ‘measure’ or the conditions of the possibility of our self-understanding and relationship to the world. See Martin Heidegger, “Science and Reflection” in *The Question Concerning Technology and Other Essays*, in which he writes:

The word *Historie* [historiography] means to explore and make visible, and therefore names a kind of representing. In contrast, the word *Geschichte* [history] means that which takes its course inasmuch as it is prepared and disposed in such and such a way, i.e., set in order and sent forth, destined. (1977 [1954]: 175)
- 2 Andrew Mitchell translates *Gestell* as positionality in “Positionality,” *Bremen and Freiburg Lectures* (Heidegger 2013).
- 3 A few disparate examples include parental intrusion into a child’s learning curve for the sake of accelerating the child’s cognitive skill set at an early age, drawing up a birth plan to optimize the woman’s birthing experience, and binge exercising.
- 4 According to Merchant (1989: 214) “Beginning with the 17th century, mechanics, “or the science of matter in motion” could be used to describe the entire universe—the human body, the physical surroundings, and the larger cosmos.” It may seem strange to challenge the assumption that matter cannot think. However, there are historical and metaphysical precedents according to which matter, or nature, is viewed as ensouled and seen to possess a kind of intelligence.
- 5 These other values—i.e., tranquility, piety, cooperation, etc.—can be seen as forming a contrast class that enables a glimpse of the historical contingency of the dominance of current technological values. On a Heideggerian reading, the recentring of the contrast class ‘values’ could signal the dawn of a posttechnological epoch.
- 6 See Dana Belu and Andrew Feenberg, “Heidegger’s Aporetic Ontology of Technology,” *Inquiry* 53, nr. 1 (2010): 1–19, for a detailed account of Heidegger’s problematic foreclosure of ontological transformation through human means.
- 7 Elective cesareans are the most enframed of all births because they express a heightened desire for order, control, and efficiency. The pregnant woman can satisfy this desire with the aid of medical technology. As is well known, elective caesareans are scheduled far in advance so the woman knows exactly when the baby will be delivered and thus she can avoid the inconvenience of unknown variables, such as waiting for labor to start, the process of pushing the baby out and feeling the pain (or pleasure) of the birthing experience. The elective cesarean is presented as a rational option, and a sign of progress. Many women think it their good fortune to have it available.
- 8 In *Choosing Tomorrow’s Children: The Ethics of Selective Reproduction* (2012), Stephen Wilkinson presents one of the standard bioethical definitions of enhancement, i.e., “The Non-Disease-Avoidance-Account” as referring to: “any improvement through modification of selection that goes beyond, or is something other than the avoidance of disease” (187–188).
- 9 One might reply that this is how ‘good’ technologies develop, through trial and error. However, a feminist retort would point out that, in this case, the cost is too high.
- 10 See Drife (2002):

The obstetric forceps have remained controversial throughout their history, and in the twentieth century a major reason was that they were used too readily and sometimes without the necessary skill. At the end of the nineteenth century and during the first decades of the twentieth, obstetrics formed a major part of general practice, and *in the interests of efficiency* a busy general practitioner would often apply the forceps rather than waiting for a normal delivery. In response to a plea for conservatism in the *British Medical Journal* of 1906, several general practitioners wrote attacking elaborate aseptic precautions as unnecessary and normal delivery as impossible for “civilised” women. This epidemic of unnecessary intervention was one of the reasons why the maternal mortality rate in Britain in 1935 was the same as it had been at the beginning of Queen Victoria’s reign. (311–315)
- 11 In 1882 German gynecologist, Max Sänger, invented the *modern* cesarean technique. His breakthrough was to suture the uterine cut after the extraction of the fetus. Before Sänger the cut was not

- closed, leaving many women to die from severe hemorrhage and sepsis. Sanger's procedure reduced maternal mortality rates. Coupled with the use of chloroform, the suturing technique increased the number of subsequent cesareans and augmented the social perception, soon to become dominant in the United States, that birthing in hospitals is safer and more progressive than home birthing. It consolidated physician-controlled birth and constituted a new frontier in obstetric intervention.
- 12 An earlier version of the spinal-epidural anesthetic was available in the early 1900s, but the lack of monitoring and skill in administration resulted in a high mortality rate. This analgesic method was abandoned for a while, landing obstetrics in what was considered to be the medical dark ages.
 - 13 This increase could be related to physicians practicing "defensive medicine." See S. Burrow (2012).
 - 14 *Reproductive enframing* applies Heidegger's concept of the enframing to the domain of reproduction. It refers to interconnected and mechanized (medical) processes during conception, pregnancy, and birth that reveal women and doctors as resources for a medical network that seeks efficiency and optimization.
 - 15 Many feminists are critical of patriarchal forces—i.e., social, political, economic, religious—sexual that influence women's reliance on reproductive technologies. Radical feminists, for example, have been most socially visible in criticizing the use of reproductive technologies. They see these as (re)new(ed), albeit more subtle, expressions of patriarchal oppression.
 - 16 In *Of Woman Born: Motherhood as Experience and Institution*, Rich warns against the patriarchal power inherent in the use of technologies that appear to help women. She cautions that "'freedom from pain', like 'sexual liberation', places a woman physically at men's disposal, though still estranged from the potentialities of her own body. While in no way altering her subjection, it can be advertised as a progressive development" (171).
 - 17 According to the most recent statistics compiled by the Centers for Disease Control, 99% of women gave birth in hospitals in the United States in 2006. Out of these, only 7.5% used midwives, while the other 91.5 % were delivered by physicians with the aid of technology. Of course, it is possible that technology was also used in midwife-assisted births but this is usually uncommon. See J. Martin, B. Hamilton, P. Sutton et al. (2009, January), Births: Final Data 2006, *National Vital Statistics Report*, 57(7), 16. www.cdc.edu. I have not been able to find statistics regarding out-of-hospital, nontechnologically mediated births.
 - 18 While the rise of the women's movement and its sociopolitical criticism of patriarchy opened the doors to natural birth, it is important to note that Lamaze in the United States (American Society for Psychoprophylaxis in Obstetrics), especially its founder Elizabeth Bing, was extremely critical of natural birth because Lamaze saw it as lacking skill and commitment. This position did not soften until the 1980s. Like natural birth, Lamaze was practiced in birth clinics. Birth clinics opened throughout the country in the early 1970s. In these lightly medicalized spaces women were attended by experienced midwives and were allowed alternatives to hospital births.
 - 19 Birth centers offer a low-tech, comfortable place for childbirth that's safer than having your baby at home if problems arise. At an accredited birth center you will be cared for by licensed professionals, usually a midwife and a nurse, with a backup hospital nearby and a doctor on call in case of an emergency. Typically, a birth center is an independent facility, though a growing number are affiliated with and often housed inside hospitals.
 - 20 Lamaze 1970 [1958]: 104–107. It is beyond the scope of this chapter to go into further physiological detail, but suffice it to say that, following Pavlov, Lamaze reinforced the connection between the cerebro-spinal system "responsible for voluntary movement and sensation" and the visceral nervous system that "governs our internal organs." Since the brain controls both systems, by controlling the information processed by the brain we can modulate the signals that the brain sends out to the rest of the body and its organs.
 - 21 Nonetheless, since Lamaze was mostly practiced in hospital settings, technological support was available if needed.
 - 22 The role of the Lamaze coach is controversial. According to some accounts, traditional coaches exerted tight authority over the birthing process, demanding obedience and even shaming women if they screamed in pain or asked for drugs to ease their pain. Today, under Lamaze International, the guidelines are much softer. See Michaels 2014: 125, 114–141.
 - 23 This slowness is also typical of birthing with a midwife, at home or in a birth clinic. Today, 'natural birth' with a midwife is a popular alternative to technophilic childbirth in a hospital. Although the midwifed woman lacks the skill and/or commitment of the Lamaze trainee (see note 19), in my view, a midwife-centered birth falls in the same technophobic category as Lamaze. Nontechnical reproductive assistance, provided by midwives, can be just as harnessed as the most sophisticated medical

machinery. It depends on how midwives are solicited and enlisted. This interesting topic remains to be explored in future work, from the point of view of the midwife and the woman who is being midwifed. For now, suffice it to say that the comparatively low skill acquisition and/or commitment on the part of the woman who births with a midwife coupled with the typical desire to optimize her birthing experience seems to situate this approach within the reproductive enframing.

- 24 After the birth of the first IVF baby (girl) in England in 1978, a large body of feminist literature emerged that criticized technological intervention in the womb as patriarchal and technology itself as patriarchal, and exposed the abuses of women by the medical industry. See Gena Corea, *The Mother Machine: Reproductive Technologies from Artificial Insemination to the Artificial Womb* (Harper & Row Publishers, 1985) and the radical feminists perspectives expressed in *Made to Order: The Myth of Reproductive and Genetic Progress* (1989), *Test Tube Women*, Rita Arditti, Renate Duelin Klein, and Shelley Minden eds. (London: Pandora Press, 1984), and Emily Martin, *The Woman in the Body: A Cultural Analysis of Reproduction* (Beacon Press, 2001).

References

- Aken, H. and Gogarten, W. (2000). A Century of Birth. *The International Anesthesia Society*, 91.
- Arditti, R., Duelin Klein, R., and Minden, S. (1989). *Made to Order: The Myth of Reproductive and Genetic Progress*. London: Pandora Press.
- Arms, S. (1994). *Immaculate Deception II*. Berkeley, CA: Celestial Arts Publishers.
- Belu, D. and Feenberg, A. (2010). Heidegger's Aporetic Ontology of Technology. *Inquiry*, 53(1), 1–19.
- Cartwright, E. (1998). The Logic of Heartbeats: Electronic Fetal Monitoring and Biomedical Constructed Birth. In R. Davis-Floyd and J. Dumit (eds.) *Cyborg Babies: From Techno-Sex to Techno Tots*. New York: Routledge Press.
- Chamberlain, D. (1998). Babies Don't Feel Pain: A Century of Denial in Medicine. In R. Davis-Flyod and J. Dumit (eds.) *Cyborg Babies: From Techno-Sex to Techno Tots*. New York: Routledge Press.
- Davis-Floyd, R. and Dumit, J. (1998). *Cyborg Babies: From Techno-Sex to Techno Tots*. New York: Routledge Press.
- Drife, J. (2002). The Start of Life: A History of Obstetrics. *Postgraduate Medical Journal*, 78, 311–315. doi:10.1136/pmj.78.919.311.
- Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*, trans. A. Sheridan. New York: Vintage Books.
- Heidegger, M. (1949). Das Ge-Stell. In Vittorio Klostermann (ed.) *Gesamtausgabe Band 79*. Frankfurt: Klostermann.
- Heidegger, M. (1977 [1954]) *The Question Concerning Technology and Other Essays*, trans. W. Lovitt. New York: Harper Publishers.
- Heidegger, M. (2012). *Bremen and Freiburg Lectures: Insight Into That Which Is and Basic Principles of Thinking*, trans. A. Mitchell. Bloomington, IN: Indiana University Press.
- Held, V. (1989). Birth and Death. *Ethics*, 99(2).
- Ihde, D. (2010). *Heidegger's Technologies: Postphenomenological Perspectives*. New York: Fordham University Press.
- Katz-Rothman, B. (2000). *Recreating Motherhood*. New Brunswick, NJ: Rutgers University Press.
- Klawiter, M. (1990). Using Arendt and Heidegger to Consider Feminist Thinking on Women and Reproductive/Infertility Technology. *Hypatia*, 5(3), 55–89.
- Lamaze, F. (1970 [1958]) *Painless Childbirth: Psychoprophylactic Method*, trans. L. R. Celestin. Chicago: Henry Regnery Co.
- Loudon, I. (2002). *Death in Childbirth: An International Study of Maternal Care and Maternal Mortality 1800–1950*. Oxford: Clarendon Press.
- Merchant, C. (1989). *The Death of Nature: Women, Ecology and the Scientific Revolution*. San Francisco, CA: Harper & Row, Publishers.
- Michaels, P. (2014). *Lamaze: An International History*. New York: Oxford University Press.
- Rich, A. (1986). *Of Woman Born: Motherhood as Experience and Institution*. New York: W.W. Norton & Co.
- Thomson, I. (2005). *Heidegger on Ontotheology: Technology and the Politics of Education*. New York: Cambridge University Press.

- Verny, T. (1981). *The Secret Life of the Unborn Child*. New York: Bantam Doubleday Dell Publishing Group.
- Wajcman, J. (1991). *Feminism Confronts Technology*. Philadelphia: Pennsylvania University Press.
- Wilkinson, S. (2012). *Choosing Tomorrow's Children: The Ethics of Selective Reproduction*. New York: Oxford University Press.
- Zwelling, E. (2001). The History of Lamaze Continues: An Interview With Sunnyie Strickland. *The Journal of Perinatal Education*, Winter, 10.

Moore's Regula

Cyrus C. M. Mody

What Does Moore's Law Say and How Does It Work?

In 1965, Gordon Moore—then the director of research for Fairchild Semiconductor—published an observation and a prediction regarding semiconductor manufacturing that has come to be known as Moore's Law. Roughly, Moore observed that the most profitable number of integrated circuit components that could be crammed onto a silicon chip (a “die”) had doubled about every year since 1959; his prediction was that this observation would continue to hold true for the next ten years. In fact, this ‘law’ has accurately described the pace of innovation in semiconductor manufacturing for almost a half century—far longer than Moore or anyone else expected.

Moore's Law has been maintained through three physical strategies (Moore 2006) and one conceptual strategy. First, until the mid-1970s, chip design was still such a rudimentary art that the refinement (and eventually the automation) of design rules freed up significant chip space for more components (resisters, diodes, capacitors, transistors, etc.). Since then, however, design rules have become so sophisticated that virtually all of the free space has been wrung out of chip architecture. Second, the number of components that can be crammed on a die has increased because the size of a standard die has increased. Die size has grown, in part, because the wafers from which dies are cut have grown—from three quarters of an inch in 1959 to 300 millimeters for standard wafers in state-of-the art semiconductor factories (fabs) today. Further increases in wafer size are still physically possible, and plans have been developed for moving to 450 millimeters. However, wafer size increases are attended by enormous capital costs for new equipment that can handle larger wafers. It's likely, therefore, that wafer size increases will be less frequent in the future; they may even cease as the returns on increasing size diminish relative to the costs of equipment replacement. Third, and most famously, the size of components has decreased steadily since 1965. In fact, some Moore's Law devotees, such as Ray Kurzweil (2005) and Hans Moravec (1999), argue that the size of components used for signal switching has been decreasing at a roughly constant rate for almost a century, if one counts quite different technologies (electromechanical switches, vacuum tubes, discrete transistors, and integrated circuits) as contributing to the same metric. That certainly is not what Gordon Moore had in mind in 1965—he was examining trends in manufacturing of silicon integrated circuits only—but it is an interesting claim worth considering. In the meantime, it is important to note that the features of some integrated circuit components in standard commercial chips are now just a few *atoms* in length. That probably means that it will become much more difficult in coming years for decreases in component size to contribute much to the furtherance of Moore's Law. Only *probably*, however, because warnings about the end of miniaturization have proven wrong for several decades. Since the search is ongoing for ways to store information at the *subatomic* scale, it is possible, if unlikely, that component size reduction will continue for many more years.

If design rules, wafer size, and component miniaturization are reaching the end of the road, what, then, will allow Moore's Law to continue? The most likely scenario under discussion in the microelectronics industry today is that the parameters of Moore's Law will have to change. Moore's Law could become a "software problem," as the same manufacturing techniques are used to make chips with multiple processors each running on their own clocks. In that case, Moore's Law would be refashioned as a measure of the steadily increasing number of operations per second performed by a commercial chip. Alternatively, the industry could figure out how to mass-produce chips that rely on a quantum computing logic (Hagar 2007). In that case, Moore's Law might be renormalized as a prediction of increase in the number of operations per second and/or number of bits stored per unit area in conventional commercial chips.

Changing criteria midstream might seem like cheating, but in fact there simply would not be a Moore's Law without continual tinkering with the metrics that define it. That is, definitional change to the law itself is the fourth strategy, this one conceptual, for furtherance of the law. As Ethan Mollick (2006) has exhaustively shown, there has never been a time when the goalposts of Moore's Law were *not* being moved. The name "Moore's Law" only became common parlance in the industry in the 1970s, after Moore left Fairchild in 1968 to start Intel, the dominant microprocessor manufacturer for the past quarter century. At virtually the same time that Moore's Law started to be referred to as such, its definition had to be modified since the rate of doubling had slowed from every year to about every two years. Further modifications have accumulated since then in order to preserve the law's accuracy. In particular, Mollick shows that the industry (especially Intel) has cannily chosen which chips to use as data points in order to make the curve of Moore's Law as smooth as possible. Thus, changing the rules midstream would only be par for the course.

Perhaps the only piece of Moore's Law that has not changed much over the years is the horizon of its predictive power. Moore originally guessed that doubling would continue at the same rate for ten years, but was agnostic as to what would come after. That is about as true today as it was then. Many knowledgeable observers do not believe Moore's Law will outlast the next decade, and even the semi-official industry definition has reduced expectations of the doubling rate for the next decade. Still, many knowledgeable observers believe that Moore's Law has been at death's door ever since the late 1960s. Readers should keep this in mind whenever they read forecasts about, or based on, extension of Moore's Law into the indefinite future. Such forecasts are based on pretty good evidence out to the ten-year mark, but are extremely hazy past that point. Inductively, the continuance of Moore's Law looks like a good bet based on its past record of proving skeptics wrong. Deductively, though, we can be pretty sure that physical laws constitute ultimate limits past which no recognizable version of Moore's Law could possibly be maintained. The insistent tug of those limits is already being felt, but when their full force will come into play is very difficult to predict.

What Kind of Thing Is Moore's Law?

In other words, Moore's Law is a law, until it isn't. In the meantime, it looks rather more like a human law, continually reinterpretable by accredited magistrates, than a natural law to which we are all bound. Like many human laws, it has a sunset provision: after ten years, its legislators revisit it. Yet what, exactly, do those 'legislators'—semiconductor industry executives; leaders and analysts in industry consortia, universities, and government agencies; trade journal editors and reporters, etc.—do? Do they decide to extend Moore's Law themselves? Do they decide whether Moore's Law is somehow extending itself? Do they decide that they are bound by its predictions? Do they simply tinker with its parameters to make the law continue to hold true?

Yes! All of these interpretations are accurate, within limits. Certainly, all of them have been put forward as explanations of Moore's Law's surprising durability. Yet under any ordinary interpretation of a 'law,' most of these views would be mutually exclusive. The key to reconciling them, I would argue, is to figure out what kind of thing Moore's Law is—and, if necessary, to invent a new category of thing into which it would fall.

First, it seems clear that Moore's Law is not a law of nature in any commonly accepted sense. There was some point—to be generous, let's say 1800—before which it is extremely unlikely that any version of Moore's Law was true; and there will quite likely be some point at which it will no longer be true. Even if the technical means to extend it were entirely understood, its continuation would be subject to all manner of seemingly extra-natural forces. If, for instance, the global economy broke down and/or the market for microprocessors tanked, there would be insufficient capital and incentive for firms to push it forward.

Yet, there is clearly some connection between Moore's Law and the laws of nature. The material behavior of silicon, silicon dioxide, aluminum, photochemical resists, developing agents, plasma and acid etches, photolithography masks, etc., all figure in the continuation of Moore's Law. Indeed, some authors—most notably Kevin Kelly (2009)—have argued that Moore's Law is immanent in the materials of microelectronics and therefore there is little we humans can do to accelerate it (and perhaps not even much we could do to slow it down).

Others take the view that Moore's Law is primarily a human construct. As David Brock puts it, "Moore's Law has not and will not happen of its own accord" (2006: 97). Obviously, no one denies that material properties figure into that construct; indeed, some of the authors most associated with this view, such as Brock (Lécuyer and Brock 2006), have forcefully argued that any historical understanding of microelectronics must give primary consideration to materiality. Rather, what authors like Brock mean is that Moore's Law holds true because the human actors involved in microelectronics manufacturing decide that it needs to hold true and work very hard to make it hold true.

This view is particularly plausible for the period since the early 1990s, when an industrial consortium, the International Technology Roadmap for Semiconductors (ITRS; previously the U.S. National Technology Roadmap for Semiconductors), began formulating roughly ten-year targets (the time horizon varies by manufacturing parameter) for innovations in the various technologies involved in processing silicon wafers. The innovation targets prescribed by the ITRS are those needed to ensure that Moore's Law continues to hold (at least, as the ITRS defines Moore's Law). The ITRS reports are filled with matrices with time on the horizontal axis and various manufacturing parameters on the vertical and benchmarks in each box. Each box is color-coded: white for benchmarks that could presently be met, yellow for ones where it is known how the benchmark would be met, and red for benchmarks that are beyond present capacity. Often (and increasingly so), the boxes to the right (i.e., further out into the future) are a sea of red; this has given rise to the colloquial saying that Moore's Law is about to hit a red 'brick wall.'

A roadmap of this sort is useful because the semiconductor industry is a fantastically complicated web of interdependent firms. No company is fully vertically integrated; instead, each firm handles just a small piece of a manufacturing process that has hundreds of different steps. Even the smallest change to any of those process steps can have significant ramifications for a dozen or more prior or following steps. No firm can make a change to its products or processes without coordinating with all of the other firms affected by that change. Changes in the manufacturing process are, therefore, incredibly difficult and have to be planned out many years in advance—and yet, such changes have to constantly be made in order to offer products that are smaller, lighter, faster, cheaper, and less energy-intensive. Continued miniaturization and higher-scale integration of circuit components make new products more appealing (consider

how often you replace your laptop) and make entirely new classes of products (cell phones, multifunction mobile devices, tablets, 'smart' glasses, etc.) feasible to market.

So some coordination mechanism is required in order to achieve the pace of innovation on which the industry and its customers have come to depend. Moore's Law offers a powerfully coordinating target (Lécuyer and Choi 2012). Every firm in the semiconductor industry, and everyone who wants to be involved with that industry (e.g., other firms, academic researchers, and government funders), can use Moore's Law as a rule of thumb to know roughly where the industry will be up to ten years from the present. They can look at the ITRS to know precisely what the industry's targets—based on Moore's Law—are for almost every conceivable aspect of the manufacturing process. Moore's Law tells every actor in and around the industry just where every other actor is headed. Actors who don't keep up with targets derived from Moore's Law get culled; actors who somehow stay a little ahead of the curve are rewarded with first-mover advantages.

It's in that sense that Moore's Law is a human construct—indeed, an actively, consciously agreed-to construct. Of course, at this point very few of the actors involved with semiconductor manufacturing actually have much say in its construction: by opting into involvement with that industry, actors agree to be *bound* by Moore's Law (or risk ejection from the industry by forces of individual- and firm-level competition), but not much more. As consumers, those of us who buy laptops, cell phones, tablets, etc., also contribute to the maintenance of Moore's Law, but again, our individual contribution is tiny.

For most of us, then, Moore's Law really does have the obduracy and objective reality commensurate with a 'law' rather than a 'construct.' Collectively, we might be able to disrupt Moore's Law, but individually our declaration that it is a fragile human construct doesn't amount to anything. There is a term from social theory and philosophy for such a law-construct—a "social fact" (Searle 1995; Giddens 1975). That is, Moore's Law is constantly made and remade by actors whose behavior is shaped by their perception that Moore's Law is a sufficiently objective fact both for themselves and for other (cooperating and competing) actors. That puts Moore's Law in much the same category as canonical social facts like paper currency, marriage, and constitutional government. Moore's Law endures partly because, like other social facts, it is malleable: what counts as Moore's Law changes from year to year, subtly enough that most participants see it as static. But Moore's Law is also durable because it is durable. That is, like other social facts, it presents an objective-enough reality for most participants that they can rely on it as an horizon for action, and in doing so endow it with that very objective reality.

Still, seeing Moore's Law as a social fact doesn't fully capture its qualities. It doesn't quite explain why the pace of doubling is what it is, and it doesn't quite take into account the push-back or "mangle" (Pickering 1995) that the material world exerts on our attempts to impose that constant rate of doubling. Of course, other social facts have some connection to the physical world, however tenuous: paper currency reflects *in some way* the scarcity of natural resources; the conventions surrounding marriage and constitutional government are tied *in some way* (if flexibly and debatably) to human biology and psychology. Yet a, perhaps small, difference is that Moore's Law was a fact before it was a social fact; Gordon Moore made his observation in 1965 of a trend that went back at least six years (if not considerably further), and that observation continued (in some sense) to hold true for many more years before it became a widely agreed-to target.

In that sense, Moore's Law more closely resembles the empirical laws governing the behavior of human collectives that are formulated by economists, political scientists, and sociologists. Economic laws, in particular, exhibit much the same mix of features as Moore's Law. They are shaped by some material obduracy in that some goods are scarcer or harder to make than others. Economic laws arise in part from the unconscious inclinations of human individuals and

collectives. But economic laws are also influenced by the conscious declarations of certain actors and institutions: presidents, chancellors of the exchequer, federal reserve boards, etc.

Indeed, it is important to note that in its original formulation Moore's Law was very much meant as an observation of a trend governed both by economics and technology. Moore observed that the cost of processing a silicon wafer was of the same order no matter how many components were on the wafer. Manufacturers therefore have an economic incentive to cram as many components as possible onto a single wafer. At any moment, current technological capacity defines a maximum number of components that can be etched onto a wafer. However, the greatest profit will always be found at some number of components *lower* than the maximum. That's because as the number of components rises, the 'yield' of functional chips (chips with components that have been correctly made) falls. Nonfunctional chips cost money to make, but they don't bring in profit because they have to be thrown away. The trick, therefore, is to cram just enough components onto the wafer that almost all of the chips are functional, but there are enough of them that you maximize sales. That's a technological rule of thumb, but it's also an economic law.

I propose that we need a new word for this kind of hybrid social-natural-technological fact. *Law* doesn't quite cut it, since Moore's Law is neither fully legislated like a human law nor objective and recalcitrant to human action as natural laws ostensibly are. Instead, I think the qualities of Moore's Law are best captured by calling it a *regula*. In English, at least, *regula* conveys multiple senses that reflect the heterogeneity of Moore's Law and phenomena like it: it is a rule to be obeyed; it is something made with a ruler, i.e., a human construct that straightens out complexity; it is a regularity observed in the world; and it has a regulatory function. Such *regula* are everywhere. In the semiconductor and computer industries alone there are at least a half dozen clustered around Moore's Law: Rock's Law, Amdahl's Law, Gene's Law, Metcalfe's Law, etc. Further afield, there are *regula* for a number of other technologies (Kim and Magee 2006)—even one (Anonymous 2006) governing growth in the number of blades that can be profitably crammed into a single razor! Sorting out the general properties and purposes of such *regula* would be a long but richly rewarding task for the philosophy of technology.

Examination of such *regula* is particularly important in the current moment because Moore's Law has entered popular and policy discourse as a model solution to all kinds of problems. For the past half-century, calls for solving social problems through a new Manhattan Project or a new Apollo program have become commonplace; the arguable success (at least on their own terms) of those technological efforts seemed translatable to solving problems such as urban decay, environmental degradation, cancer, energy independence, etc. Today, though, calls for a 'Moore's Law' for climate change remediation or a 'Moore's Law' for solar energy have begun to occupy the niche in at least American discourse that used to belong to gestures to the Manhattan Project or the moon race. If the seeming inevitability and speed of innovation governed by Moore's *Regula* could be translated to, say, year-on-year improvements in the efficiency and cost of solar cells, well, fossil fuel use would drop precipitously very soon.

So far, though, the Moore's *Regula* template has stubbornly resisted transfer to other domains. That's not for lack of trying. In particular, as Ann Johnson (Johnson and Stone 2010) has shown, policy makers in government and industry have become enamored in the past decade or so of roadmaps modeled on the ITRS. The idea seems to be that if all the relevant actors associated with some technology or problem area just agree to hit a road mapped set of benchmarks, then those benchmarks will magically get met—as the semiconductor industry appears to have done so successfully. Of course, the benchmarks that the semiconductor industry hits first arose more or less organically and were only gradually turned into explicit horizons of action. Even then, those benchmarks simply could not be met without the full weight of the hundreds of firms, universities, and government agencies—not to mention vast market demand—that

undergird microelectronics innovation. That level of activity and economic incentive has so far not emerged for the regula imposed in fields such as solar energy, and so the roadmaps for innovation in those fields have either been trivial or unrealistic. Philosophical reflection might, however, offer a better sense of whether these kinds of regula can be commanded into existence (in which case, how?) or must arise organically (in which case, maybe the roadmap frenzy can abate somewhat).

What Does Moore's Regula Mean for Philosophy of Technology?

The institutional and economic infrastructure supporting Moore's Regula exists because of the extraordinary economic, social, and even legal importance of microelectronics innovation—an importance that invites significant philosophical inquiry. Electronics is the manufacturing industry that contributes more to U.S. GDP than any other; because of the globalization of that industry, much the same could also be said of many other countries, particularly in East and Southeast Asia (among others: Japan, South Korea, Taiwan, Malaysia, and increasingly the People's Republic of China). Many ancillary industries are only possible, or at least economically feasible, because of the extreme speeds and light weights of integrated circuits that Moore's Regula has afforded (Predicted? Provided? Enabled?). Mobile apps, cloud computing, outsourced call centers, telecommuting, desktop publishing—these all would not exist or would not look much like they do were integrated circuit miniaturization not governed by Moore's Regula.

That means that there are many more things and practices in the world than there would be without Moore's Regula. Semiconductor manufacturing has become, in Paul Rabinow and Talia Dan-Cohen's (2005) phrase, “a machine to make a future”—i.e., a technology from which new human possibilities and new grounds for ethical action and responsibility open up. Oddly, Rabinow and other Foucauldian authors have focused almost entirely on the life sciences and ‘biopolitics,’ but it seems likely that the new ethical horizons opened up thanks to Moore's Regula will be as profound as those that follow from genomics or recombinant DNA. Indeed, genomics technology only has the future-making potential that it does because biotechnologists and electrical engineers have developed ‘gene chips,’ automated sequencers, and Big Data statistical algorithms for understanding and manipulating genetic material: all of these techniques require the fast computational speeds achieved via Moore's Regula.

Yet Moore's Regula has been, is, and will be woven into new future-making technologies and practices far beyond biotechnology. For instance, the technologies of modern surveillance—storage of conversational metadata, real-time ability to monitor virtually any phone conversation, ubiquitous cameras linked to databases for face recognition, etc.—that the U.S. National Security Agency and many other state and nonstate actors employ would not be feasible without fast, small, cheap integrated circuits (Weckert 2002). The same goes for the aerial drones that the U.S. military and intelligence services currently operate and that many other state and nonstate actors either currently or will soon have access to. Rudimentary ‘drones’ have been around since before World War II (Chandler 2014), but they are a hot-button ethical and political issue today because microelectronics has reached the point where advanced drones can be operated in real time from the opposite side of the world, can feed a variety of different information about their environments back to their operators, and can even have a certain amount of autonomous decision making.

Indeed, the increasing autonomy of electronic agents is probably Moore's Law's thorniest gift to philosophers. Increasing circuit speed has brought a host of new technologies that seem very close to artificial intelligences (AIs). Siri may not quite be able to pass the Turing Test, but

‘she’ can uphold her end of a conversation about as well as plenty of humans. Google Translate doesn’t perfectly convey the meaning of text on the websites of my favorite Korean soap operas, but then the humans who do the subtitles for those soap operas don’t get everything right either.

That is, we’re not yet at the point where we think of software as ‘intelligent’ in the same way we do a human, but we have arrived at a point where more and more activities that only humans used to do will now be done by electronic agents. They won’t do those activities perfectly, but they’re now doing them well enough that the ordinary mechanisms of ‘repair’ (Schegloff 1992) that humans continually apply to each other’s words and actions will now be sufficient to make machines’ words and actions intelligible as well. Even if Siri is a lousy conversationalist, she’s good enough that I can ‘get’ what she means using the same level of repair that I would use to ‘get’ what my friends mean when they too are being lousy conversationalists.

Curiously, though, most of the new machine ‘intelligences’ that have become cultural phenomena in the past decade (as well as the ones that invisibly run much of the world’s infrastructure) don’t operate on the principles that the pioneers of artificial intelligence thought they would in the 1950s. Back then, artificial intelligence research looked a whole lot like applied philosophy of science. The field’s leaders, such as Herbert Simon, viewed science as the exemplar of humanity’s distinctive rationality and intelligence, and therefore strove to get machines to operate the way they thought scientists did. That line of work produced some interesting results, but it seems fair to say that it did not succeed on its own terms (Roland and Shiman 2002; Collins 1990). Scientific, and general human, reasoning are a lot more complicated than those pioneers thought. Scientists, like other humans, have bodies, are social beings, operate in situational contexts with their own local logics, etc. As Hubert Dreyfus (1972; 1992) argued, there are plenty of things that a philosophy of science-based AI can’t do, and the most sophisticated things they can do (e.g., beat a chess grandmaster) don’t look that translatable to other human activities.

Yet that’s not how the current crop of AIs work. Siri can’t discover Newton’s Laws, but she can compare your words to snippets of speech in her files and then look up what phrase she could acceptably say in response. She’s a huge database with an extraordinarily fast searching algorithm—not a scientist. Similarly, many of Google’s algorithms operate on weighted rules of thumb, rather than on ironclad highly rational deduction. Perhaps the clearest example of the difference between the old and new approaches can be seen by comparing IBM’s chess champion, Deep Blue, to its Jeopardy champion, Watson. Built in the late 1990s, Deep Blue was incredibly ‘smart,’ but it knew very little. The Watson of today is kind of bumbling, but it knows a lot, and it’s just smart enough to figure out which thing that it knows is relevant in the current moment.

That means that the grounds for debate in the philosophy of artificial intelligence are changing thanks to Moore’s Regula. The old debate was primarily about whether machine *logic* was sufficient for machine *intelligence*. Today’s fast, cheap, high-density circuits, however, mean that machine speed, storage, and pattern recognition are (almost?) as important as logic. The computers that Dreyfus said “can’t do” lots of human activities were slow and couldn’t remember much. But what things can or can’t computers do now that they have near- or super-human storage capacity and searching/processing/pattern-matching speed? Dreyfus’s basic point—that computers are phenomenologically different from humans and inhabit the world in a different way than we do—still seems safe, but some of the details of his argument will likely need to be amended.

One speculative possibility is that the increasing autonomy of electronic agents has brought us close to the point where such agents will be able to do philosophy for themselves. Ray Kurzweil (2005) has famously referred to that point as “the Singularity”—the moment at which the

density of logic and memory on an integrated circuit surpasses the density of logic and memory in the human brain. At that moment, Kurzweil argues, computers will be able to do most human tasks better than we can—including, crucially, the task of innovating the design of computer software and hardware. Once that happens, electronic agents will become exponentially smarter than human agents more or less overnight. As a result, human existence will be transformed completely and in very short order—if, indeed, humans will continue to exist at all.

Talk about the Singularity has infected techno-utopian discourse—particularly in Silicon Valley—to an unaccountable extent in the past decade. There is even now a Singularity University! Yet, there's still plenty of room for critical analysis of what the Singularity is, how likely it might be, or whether we would even notice that it had happened. I would suggest, therefore, that there is a good philosophical study of the Singularity in the offing, one that could pick apart the assumptions implicit in Kurzweil's specific vision of radical technological transformation but that could also illuminate some of the characteristics of techno-transformative visions in general.

One particular assumption of the Singularitarians is that it *will* happen. Humans might be able to move the date of the Singularity forward or backward slightly, and we might have a little say over whether it is overall a good or bad thing for the human species, but to a large extent we are simply being borne toward this epochal break. In other words, Singularitarianism rests on an implicit or even explicit technological determinism (Mody 2004). Indeed, Singularitarianism incorporates multiple kinds of technological determinism: e.g., the belief that technology has an internal, immutable logic; that technology is an active force that can transform societies; that the level of technological “achievement” is the yardstick by which societies should be measured and compared; and so on.

Technological determinism has a long history as a subject of philosophical inquiry, going back at least to Marx, and so the Singularitarian version could simply be taken as yet another case for exploration of that topic. Yet there is one aspect of the Singularitarians' determinism that offers grounds for a new analysis: Moore's Regula. If Moore's Regula continues to hold, even at some reduced rate of doubling or under some modestly altered definition, then we will—more or less inevitably—move very rapidly toward extremely capable electronic agents. It's still up for debate whether that means we will inevitably arrive at the Singularity, but we will at the least be on a trajectory toward realization of some of the conditions Singularitarians have identified as necessary for their epoch-making event. That's because Moore's Regula seems, at least on initial inspection, to be a deterministic law. If technological determinism has any plausibility to it at all, Moore's Regula is probably the best evidence we have for it (Ceruzzi 2005). Yet accepting the inevitability of Moore's Regula is itself a contingent act with some unsavory consequences. For instance, if Moore's Regula is a deterministic law, does that absolve its stakeholders (i.e., virtually everyone) from responsibility for the environmental degradation brought about by semiconductor manufacturing, cloud computing, and other extremely water- and energy-intensive applications of advanced microelectronics (Patrignani, Laaksoharju, Kavathatzopoulos 2011)?

In the meantime, on our way to the alleged Singularity, the consequences of Moore's Regula are being felt in ways that my undergraduate students believe to be societally transformative. Perhaps the best way to think about the present sociotechnical moment is that we have reached the ‘postindustrial’ era predicted by Daniel Bell (1973). There is, of course, plenty of old-line industry still going on; much of it today has moved to parts of the world that were still ‘industrializing’ when Bell was writing, but there's still a great deal of manufacturing done in the long-industrialized countries as well. ‘Thanks’ (whatever that means) to Moore's Regula, however, more and more manufacturing is done by fewer and fewer people, with enormous consequences for global political economy, the relationship between labor and capital, the state of the working and middle classes (as well as the class of the ‘1%’), and so on.

Ironically enough, many of the old-line manufacturing firms that made Moore's Regula possible have contracted in the postindustrial era it has brought about. RCA, Bell, and Fairchild Semiconductor—three of the pioneers of transistor and integrated circuit technology—no longer exist. Many other vertically integrated giants of microelectronics have become much smaller and more specialized in the past twenty years: IBM, Hewlett-Packard, Texas Instruments, etc. These companies were once *the* models for the multidivision corporations that dominated the economies of North America and Europe (Chandler 1977).

In their place, the products associated with Moore's Law have opened up new forms of enterprise, broadly construed. The 'financialization' of many postindustrial economies, for instance, would be infeasible or at least unprofitable if not for high-frequency trading servers, automated trading 'bots,' and massively networked flows of capital. More broadly, the effects of Moore's Regula can be seen in the decentering and blurring of consumption, production, and innovation in the postindustrial era. It used to be that most countries were served only by a handful of television networks (often state-owned), a state-owned telephone monopoly, an oligopoly of firms that dominated manufacturing and R&D, and a population that made do with the products offered by those giants. Innovation of course still went on all over the place—a lot of individual users' ingenuity went into that 'making do,' especially since manufacturing giants weren't all that attuned to the needs of nonaffluent and especially non-Western users (Edgerton 2007). But the distribution of innovation seems to have been quite different in that industrial era than it is today. Now, many more people than before can create a software program, a television program, a radio show, a newspaper column, or even a physical product, and distribute it for very wide circulation. We might call these things by different names than we did before—a 'blog' rather than a 'column,' a 'YouTube video' rather than a 'television show'—but the analogs are relatively clear.

That should be an invitation, I think, for philosophers of technology to ponder how some of the basic concepts associated with technology might be changing as the technologies associated with Moore's Regula become faster, cheaper, and more ubiquitous. It seems to me that we may need a new understanding of innovation, invention, production, consumption, manufacturing, and even technology itself in an era where fast, cheap circuits mean that cultural products such as video games and movies are as technologically intensive and as important to postindustrial economies as manufactured products are, and where 'high-tech' manufactured products are likely to be made by hand or with a three-dimensional printer and distributed through networked communities of users and producers. In other words, philosophy of technology needs to be flexible enough to deal with a world where consumption and production are governed by Etsy and Facebook rather than Siemens and DuPont. Even if Etsy and Facebook are only around for a short time, they demonstrate the *possibility* of a certain mode of innovation, consumption, and production that should be philosophically accountable to the same extent that more classically 'industrial' and 'preindustrial' forms are.

References

- Anonymous. (2006). The Cutting Edge: A Moore's Law for Razor Blades. *The Economist*, March 16.
- Bell, D. (1973). *The Coming of Post-Industrial Society: A Venture in Social Forecasting*. New York: Basic Books.
- Brock, D. C. (ed.) (2006). *Understanding Moore's Law: Four Decades of Innovation*. Philadelphia, PA: Chemical Heritage Foundation.
- Ceruzzi, P. E. (2005). Moore's Law and Technological Determinism: Reflections on the History of Technology. *Technology and Culture*, 46, 584–594.
- Chandler, A. D. (1977). *The Visible Hand*. Cambridge, MA: Harvard University Press.

- Chandler, K. (2014). *Drone Flight and Failure: The United States' Secret Trials, Experiments, and Operations in Unmanning, 1936–1973*. Ph.D. diss., Berkeley.
- Collins, H. M. (1990). *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge, MA: MIT Press.
- Dreyfus, H. L. (1972). *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper and Row.
- Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.
- Edgerton, D. (2007). *The Shock of the Old: Technology and Global History Since 1900*. New York: Oxford University Press.
- Giddens, A. (1975). *New Rules of Sociological Method: A Positive Critique of Interpretive Sociologies*. New York: Basic Books.
- Hagar, A. (2007). Quantum Algorithms: Philosophical Lessons. *Minds and Machines*, 17(2), 233–247.
- Johnson, A., and Stone, J. (2010). Roadmaps. in D. H. Guston (ed.) *Encyclopedia of Nanoscience and Society*, vol. 2. Thousand Oaks, CA: Sage.
- Kelly, K. (2009). Was Moore's Law Inevitable? *The Technium blog*, July 17. Available at: www.kk.org/thetechnium/archives/2009/07/was_moores_law.php [Accessed November 29, 2012].
- Kim, H., and Magee, C. L. (2006). A Functional Approach for Studying Technological Progress: Application to Information Technology. *Technological Forecasting and Social Change*, 73, 1061–1083.
- Kurzweil, R. (2005). *The Singularity Is Near*. New York: Penguin.
- Lécuyer, C., and Brock, D. C. (2006). The Materiality of Microelectronics. *History and Technology*, 22, 301–325.
- Lécuyer, C., and Choi, H. (2012). Silicon Valley Secrets, or U.S. Microelectronics Firms Faced With Technical Uncertainty. *Revue d'histoire moderne et contemporaine*, 3, 48–69.
- Mody, C. C. M. (2004). Small, But Determined: Technological Determinism in Nanoscience. *Hyle*, 10, 99–128.
- Mollick, E. (2006). Establishing Moore's Law. *IEEE Annals of the History of Computing*, 28, 62–75.
- Moore, G. E. (1965). Cramming More Components Onto Integrated Circuits. *Electronics*, 38(8), 114–117.
- Moore, G. E. (2006). Moore's Law at 40. In D. C. Brock (ed.) *Understanding Moore's Law: Four Decades of Innovation*. Philadelphia, PA: Chemical Heritage Foundation.
- Moravec, H. (1999). *ROBOT: Mere Machine to Transcendent Mind*. New York: Oxford University Press.
- Patrignani, N., Laaksoharju, M., and Kavathatzopoulos, I. (2011). Challenging the Pursuit of Moore's Law: ICT Sustainability in the Cloud Computing Era. *Noticia di Politeia*, 27, 45–55.
- Pickering, A. (1995). *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.
- Rabinow, P., and Dan-Cohen, T. (2005). *A Machine to Make a Future: Biotech Chronicles*. Princeton, NJ: Princeton University Press.
- Roland, A., and Shiman, P. (2002). *Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983–1993*. Cambridge, MA: MIT Press.
- Schegloff, E. A. (1992). Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *American Journal of Sociology*, 97, 1295–1345.
- Searle, J. (1995). *The Construction of Social Reality*. New York: Simon and Schuster.
- Weckert, J. (2002). Lilliputian Computer Ethics. *Metaphilosophy*, 33, 366–375.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Inner Space and Environment



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Is Renewable Energy Technology Revolutionary?

Robert Darrow

Four decades ago, a motley construction crew assembled at an alternative high school campus on the outskirts of the railway town of Ulfborg, in rural western Denmark, and built what was at the time the largest wind turbine in the world.

For three years the Tvind School's windmill team planned and worked by committee, with design assistance from several technical university professors, but with no prior technical training or fabrication experience. New manufacturing techniques were invented through trial and error, and trade skills were learned on the job. Construction was largely financed with the shared salaries of the Tvind faculty, who were members of the Teachers Group, a communist collective many Danes suspected of being a cult. Students drawn by newspaper advertisements traveled from around the country to contribute their unpaid labor, working long hours in conditions that some sources have described as "Maoist" in their harsh discipline. Tapping into growing environmental consciousness in Denmark in the wake of the 1973 oil crisis, the windmill project aimed to protest the then-current government's plans to build several nuclear reactors, and to demonstrate that ordinary people could supply their own clean, sustainable energy.

Forty years later, the Tvindmill is still feeding electricity to the grid. When the working group reunited in early summer 2015 to celebrate the anniversary of their accomplishment, one former student worker spoke about the hundreds of modern turbines she had passed driving along the highway from Copenhagen that morning, and the ensuing realization that their efforts had helped launch a renewable energy revolution.

It is hard to deny the influence this ragtag collective of leftist activists and amateur technologists had in spurring an energy transition that is now passing the tipping point in Denmark. Their work on the 'Tvindmill' inspired and helped organize a generation of wind energy pioneers, who in the late 1970s gave birth to the modern wind industry in the Danish countryside. Soon after, the government's proposed nuclear program was abandoned, largely as a result of pressure from grassroots opposition groups. Today, Denmark leads the world in wind energy production per capita, generating about 40% of its electricity from wind in 2014. The national government recently set an ambitious target, with overwhelming public support, of generating 100% of the country's energy needs from renewable sources by 2050. No other industrialized nation has made so much progress in replacing fossil fuels with renewable energy sources.

But in the intervening decades, Denmark has moved on from the Tvind model of development. No longer is wind energy the domain of amateur enthusiasts, patching together prototype turbines from spare truck parts and secondhand factory equipment. The Danish wind sector has grown into a global industry dominated by publicly traded manufacturing giants, national utilities, major investment banks, and professional engineering firms collaborating on multi-billion dollar projects. The distinctive culture of cooperative development and community self-sufficiency that incubated early entrepreneurs has been pushed to the margins. To some, the scaling up of the industry is proof that wind power has finally arrived and is here to

stay. Convincing the big utilities and big banks to invest means wind turbines are no longer a fringe or niche technology: they have gone mainstream and now have the potential to supply a substantial share of Denmark's, and the world's, energy needs.

For the idealists, though, assimilation was never the goal. They view this trend with regret and are nostalgic for a more democratic, human-scaled approach to energy production. The unique characteristics of renewable energy brought hope for a dramatic transformation of everyday life. It is not readily apparent that such a transformation is occurring. The air in Copenhagen is cleaner than it used to be, and more of the economic benefits of energy production are staying at home. Overall energy consumption has actually decreased since the 1970s, in spite of continued GDP growth, although those energy savings have been mostly achieved through efficiency gains, such as better-insulated buildings and the expansion of district heating systems. These are all significant changes and improvements in Danish society. But for most Danes, the rhythms of daily life have been relatively unaffected by the widespread adoption of wind energy. They have not sacrificed any modern conveniences, or radically curbed consumption patterns. They still drive gasoline-powered cars, still eat processed foods, still surf the web, and shop at the mall. Denmark today is modern, capitalist, and industrialized, and in broad strokes, so is its wind business.

Were the seismic shifts in Denmark's energy landscape merely the temporary rumblings of a dominant culture adjusting to, and absorbing, a new entrant on the scene? Or are the Danes in the midst of an unfolding revolution?

For as central a concept as revolution has been to the history of technology, it remains a relatively undertheorized one. When Larry Hickman wrote about "paths to technological revolution" in the first volume of *Techné*, he offered little description of the revolution he sought, only that it would politicize technology and, following John Dewey, democratize technology (Hickman 1996). Nick Bostrom had similar difficulty pinning down the concept, claiming that technological revolutions are "among the most important things that happen to humanity," but admitting that his definition of revolution as "a dramatic change brought about relatively quickly," was "rather vague," and relying instead on illustrations such as the invention of agriculture, the printing press, and the atomic bomb (Bostrom 2007: 1–2). Though technological revolutions are typically associated with new innovations—as in Bostrom's examples—there seems to be agreement that a mere technical development is not sufficient for a revolution. A technological revolution describes a more fundamental change in human societies, capacities, and understanding. Thus, when Peter Drucker wrote about the technological revolution of the eighteenth century, he described it as "a major change in an attitude, a world view" (Drucker 1961: 344).

Hannah Arendt, who has thought more deeply about the meaning of revolution than most, argues that the modern usage of the term arose in a specific historical context. The original Latin meaning of *revolution*, employed by Copernicus, referred to regular, cyclical motion. When the term was first used to describe political events during the Glorious Revolution in England, it meant the restoration of the former order. It was only during the period of the American and French revolutions that the concept developed its modern associations with violent rupture, the alteration of the course of history, and the appearance of something new under the sun (Arendt 1963: 21–40).

Wind turbines were a novel, disruptive means of generating electricity in the 1970s, and the Tvindmill team certainly saw their construction project as part of a larger struggle against the monopolization of the energy system that would return power to the people. Many in the tightly knit Danish wind community of that period shared similar ideological motivations for pursuing renewable energy projects, and sympathized with the 'small is beautiful' vision of the appropriate technology movement. The history of modern wind power in Denmark, with its politicized

origins and unrivaled pace of development, offers a unique case for evaluating current theories of large-scale technological change and testing common assumptions about the transformative potential of renewable technologies by studying an energy transition in progress.

Renewable energy has long been associated with green politics and the hope that a society no longer reliant on fossil fuels would also be more democratic, more egalitarian, more respectful of its environment and its limits, and a better steward of natural resources. This view was perhaps best encapsulated in Amory Lovins's influential "hard and soft energy paths" framework, which contrasted the centralized and authoritarian characteristics of "hard" nuclear and fossil energy technologies with the "soft path" offered by renewable energy technologies, which Lovins characterized as "flexible, resilient, sustainable, and benign" (Lovins 1977: 38).

It would be unfair to ascribe to Lovins the belief that the adoption of green energy technology would automatically usher in a utopian society, though perhaps some of his disciples held this hope. Lovins himself maintained that pursuing a soft energy path entailed not only technical fixes, but also corresponding changes in the economy, society, politics, and values. He wrote that the "most important, difficult, and neglected questions of energy strategy are not mainly technical or economic but rather social and ethical" (Lovins 1977: 58). Still, it is clear that he saw soft and hard technological paths as deeply coupled with particular political values and modes of social organization. Langdon Winner has frequently been charged with determinism for proposing a category of "inherently political technologies." He has expressed particular concern about the authoritarian structures he believes are necessary to support nuclear power, a worry he and Lovins share. Though both seem comfortable characterizing the hard paths of nuclear and fossil fuels as inherently undemocratic, they are less willing to claim the opposite for alternative fuel sources. "The social consequences of building renewable energy systems," Winner writes, "will surely depend on the specific configurations of both hardware and the social institutions created to bring that energy to us" (Winner 1986: 39). Writing within this framework of uncertainty about how renewable energy systems will develop, Sujatha Raman has recently argued that current practices in the alternative energy sector more closely resemble a hard path, and that renewable technologies are becoming "fossilized" through the political economies of their supply chains and production processes, such as the use of rare earth metals in wind turbine components. It is notable that Raman's analysis emphasizes the material characteristics of the technical artifacts themselves, but her broader thesis that renewable energy systems are starting "to resemble the fossil fuel regime they are supposed to supersede," raises the disquieting notion that the opportunity to build a culture of soft energy production and consumption based on alternative fuels may be receding into the past, the road not taken (Raman 2013: 172).

The ambitions of the environmental movement of the 1960s and 1970s may seem naively optimistic in retrospect and suggestive of a deterministic outlook no longer in vogue among philosophers of technology. But the question of the relationship between technological change and broader sociopolitical change remains critically important in a world facing a climate crisis and fast exhausting its fossil fuel reserves.

Some scholars argue that the transition to a sustainable energy system can be achieved with limited disruption to industrialized societies. Ecological modernization theorists propose that environmental problems can be solved with technological advances, and the pursuit of such solutions is an opportunity for economic growth. Similarly, advocates for the 'greening of capitalism' suggest that the dominant economic system can be reformed, through the creation of tools like carbon trading schemes and the development of green consumer products, to create market incentives for more environmentally responsible behavior (see, for example, Mathews 2015). On an optimistic assessment of the potential of market-based solutions, renewable energy need not be revolutionary in anything more than a purely technical sense. Advanced

societies can simply phase out dirty fossil fuel technologies for better, cleaner methods of generating watts and carry on with business as usual. A dimmer view emphasizes resilience in the face of inevitable shifts in climate conditions and resource availability, and proposes market systems as the most nimble and efficient means of adapting to emerging challenges and opportunities.

But such orthodoxy-preserving solutions have been challenged by numerous commentators who see capitalism as the chief culprit of environmental destruction, rather than a potential savior. Among these critics are those who claim there is an inherent conflict between the expansionary logic of capitalism—what sociologist Allan Schnaiberg has termed the “treadmill of production”—and the steady-state economies necessary for sustainable living. Inspired by Marxist economic science, adherents of this view contend the ever-present need for accumulation in capitalist societies will run up against the limits of natural systems, and the capitalist mode of production will collapse under the weight of its own internal contradictions. A related argument holds that industrial capitalism depends on the extraction of value from human labor and natural resources, which puts it fundamentally at odds with a worldview that emphasizes respect and care for all forms of life. It may be possible to devise market mechanisms for controlling emissions of CFCs or atmospheric carbon, but human and nonhuman populations with no immediate monetary value will continue to be ignored at best, and exploited at worst. The unequal distribution of costs and benefits produced by such piecemeal solutions represents not only a basic injustice, but also a recipe for dangerous unintended consequences. Appealing to the precautionary principle, this view suggests that engineering stopgap responses to environmental hazards targets the symptoms rather than the disease, and has the potential to produce side effects that are as bad or worse than the problems they were intended to solve. Environmental degradation cannot be alleviated without first addressing its underlying social and political-economic causes. Thus, a just and sustainable solution to environmental crisis requires a broader cultural shift away from what author Naomi Klein has termed the ideology of “extractivism.” Sustainable energy use, she writes, demands “we adapt ourselves to the rhythms of natural systems, as opposed to bending those systems to our will with brute force engineering” (Klein 2014: 394).

Klein recognizes the possibility that future societies could find ways to address the challenge of climate change without a radical overhaul of the status quo social order and its dominant ideology. But whereas this scenario represents the ideal outcome for ecological modernists, Klein and scholars concerned with environmental justice would view such a path forward as a missed opportunity. “The shift from one power system to another must be more than a mere flipping of a switch,” Klein writes, “It must be accompanied by a power correction in which the old injustices that plague our societies are righted once and for all” (Klein 2014: 399). It may very well be the case that the adoption of a particular energy regime is not closely linked with any corresponding mode of social life, but environmental justice advocates see the period of transition as a moment when broader social changes can be considered and perhaps catalyzed—a moment with revolutionary potential.

Four Myths About Renewable Energy

When Amory Lovins coined the terms “hard and soft energy paths” in the 1970s, he described the soft path as (1) reliant “on renewable energy flows,” (2) “diverse,” (3) “flexible and relatively low technology,” (4) “matched in *scale* and in geographic distribution to end use needs,” and (5) “matched in *energy quality* to end-use needs” (Lovins 1977: 39). Certain, supposedly inherent, technical properties of wind turbines and solar panels have long been assumed to exemplify these qualities. Thus, renewable technologies have been associated with democratic,

decentralized, and environmentally friendly energy production, in contrast to the “hard” path of the authoritarian, polluting, unsustainable fossil fuel and nuclear industries. Renewable energy has frequently been championed by activists aiming to remake society, and summarily dismissed by big energy firms as unreliable, inefficient, and unrealistic.

Lovins never suggested that the distinction between the hard and soft paths was merely technical, and he held that the “sociopolitical *structure* of the energy system” was equally important. (Lovins 1977: 38) The Danish experience with wind energy demonstrates that the context in which the technology is adopted, and the purposes for which it is employed, have a far greater influence in shaping the energy system than any technical characteristics of the turbines themselves. Wind turbines have proven to be an extremely flexible technology, suitable for a wide range of applications. The following sections briefly examine four common assumptions about renewable energy that have been called into question by recent experiences with wind power in Denmark.

Renewable Energy Is Intermittent

The wind is a free and seemingly inexhaustible resource, created by the sun’s uneven heating of the Earth’s surface—but the wind does not blow in all locations at all times. (Solar power poses an equivalent challenge: the sun does not always shine.) For electricity grid managers, this intermittency is a problem, because electric demand follows fairly predictable cycles, which do not necessarily correlate with the supply of wind. This unreliability of supply has led detractors to claim that wind turbines can never provide more than a small share of a society’s energy needs, since backup generators must always be ready to compensate for lulls in the wind.

The Danes are proving that such pessimistic assessments show a distinct lack of imagination. Within the next few years, wind turbines are expected to supply more than half of the nation’s electricity. This feat has been achieved without having to build new fossil-fueled backup capacity through more efficient grid management and expanding the grid to neighboring countries. The entire Scandinavian grid has already been integrated with cables from Denmark to Sweden, Norway, and Germany. The linchpin of this arrangement is Norway’s vast hydroelectric resources, which form a symbiotic relationship with wind power from Denmark and Germany. When Denmark produces more wind energy than it can use (now a frequent occurrence on windy days), the excess is exported to Norway at discounted rates. In this way, Norway is able to hold hydroelectric capacity—which is easy to stop and start at a moment’s notice—in reserve, and open its dams on less windy days when the market offers better prices. The integrated grid and electricity market allows the available supply at any given moment to be dispatched across a wide geographic area. Additional undersea cables are already being planned to connect with Germany, the Netherlands, and eventually the United Kingdom. The long-term hope is to connect all of Europe in one integrated grid system.

There are other possible means of addressing the challenge of intermittent renewable energy supply, such as storing excess electricity production during windy periods or adjusting patterns of energy use to correspond with the availability of wind power. As the smart grid is slowly being implemented in Denmark, there are proposals to create price signals that would encourage consumers to shift energy-intensive activities such as washing clothes and charging electric vehicles, and even industrial production, to periods of lower energy demand. Next-generation batteries and other storage technologies currently under development may make household electricity storage feasible in the coming decades, and the dream of going off the grid without sacrificing the convenience of on-demand power may finally be realized. Danish power companies have also begun investing in large electric boilers that would make it possible to store wind energy at central co-generation plants in the form of hot water.

Renewable Energy Is Distributed

Because renewable resources are geographically dispersed and cannot be easily stored and transported the way coal and oil can be, it has been thought that renewable energy would have to be consumed near the site where it was produced. This constraint would presumably lead to a highly distributed energy network. The opposite has occurred in Denmark, where a growing share of renewable sources in the energy mix has corresponded with increasing centralization of the energy system. Today, Denmark's entire electric grid is managed from a single computerized control room at the national transmission system operator's headquarters in the town of Fredericia. Until a 2005 merger created the state-run Transmission System Operator, Energinet.dk, responsibility for the Danish grid had been divided among numerous regional utilities, often cooperatively or municipally owned. Energinet.dk's parent, the Danish Energy Agency, is involved in almost every aspect of wind power development, from long-term planning to environmental review of individual projects to building the grid infrastructure necessary for connecting new turbines to, crucially, administering the electricity tax receipts that incentivize renewable energy development.

The traditional model of wind development in Denmark, in which a single turbine or a small cluster of turbines is erected, usually by a landowner or a locally based cooperative, is being replaced by fewer, larger arrays of turbines, and increasingly by offshore parks. While production at offshore wind parks can benefit from higher and steadier winds at sea, their cost of energy is still far higher than building onshore, due to enormous construction and maintenance costs. A distributed network of small producers would arguably pose lower investment risks and would be more resilient, given the possibility of outages at centralized power plants. Still, the utility companies prefer to build massive offshore projects, largely for organizational reasons. It's easier for a company to devote staff resources to pursuing one large project than to juggle a portfolio of smaller projects. Denmark's largest utility recently announced it would divest from its onshore assets entirely. Government policy has also supported this trend, offering tenders for several new offshore wind parks in just the last few years.

Renewable Energy Is Appropriately Scaled

Until the mid-1980s, the only commercially available wind turbines in Denmark were tiny by today's standards, with a maximum power output of 10kW to 25kW. They were affordable and relatively easy to install, and could often be maintained by their owners. Most of these turbines were purchased by families looking to replace their household consumption in rural settings, where energy costs tended to be higher. These early turbines exemplified the concept of appropriate technology, most closely associated with the writings of E. F. Schumacher, which emphasized small-scale, low-skill technologies that matched the needs of the community of users and blended harmoniously with their surroundings.

The principles of appropriate technology were left behind as wind turbines shifted from being an ideologically motivated experiment to an industrial concern in the 1980s. Turbines grew iteratively in size, from 225kW to 500kW and then 1MW, but most installations remained small-scale, consisting of only a single turbine or a small cluster of turbines, owned by private individuals or community cooperatives. This ownership model was supported by a national law that required owners to live within a certain distance of their turbines. This restriction was lifted in the mid-1990s, with dramatic consequences for onshore wind development. Generous government subsidies had long made wind turbines a good investment, but the local ownership requirement had kept speculation to a minimum. The new subsidy schemes, as well as new municipal

planning rules, allowed private developers to rush in and buy up the most favorable locations for new installations. The cooperative model of development began a steady decline, replaced by for-profit firms owned by out-of-towners. This change had several other unintended consequences. First, unlike local cooperatives, these private developers were not interested primarily in replacing local electricity consumption with a clean energy source, but in maximizing the return on their investment. So they requested ever larger turbines from the manufacturers, and the average size onshore grew to 2 MW, and in recent years to 3.6 MW. These mammoth modern wind turbines have a far greater impact—in terms of both noise and visual disturbance—on neighbors. Whereas the turbines of the 1980s rose 20 or 30 meters above the ground, today's onshore turbines reach a height of up to 150 meters—the size of a 40-story skyscraper. Furthermore, locals were no longer receiving direct economic benefits from these projects. Not surprisingly, local resistance to new turbine developments has risen sharply in the last two decades.

Renewable Energy Is Highly Visible

There is arguably no energy technology in use today that is more visible than wind power. In Denmark, it is hard to go anywhere without seeing turbines spinning on the horizon. Since the wind is best at higher elevations and where there are few obstructions, and since the eye is drawn to the spinning motion, turbines are exceedingly difficult to camouflage. Although many cases in Denmark, and elsewhere around the globe, show that locals are often extremely supportive of wind power when they see direct local benefits, and farming communities are usually more accepting of the visual impacts than vacationers, it is reasonable to assume there is a correlation between the growing scale and penetration of wind turbines and the not-in-my-backyard opposition to the onshore industry.

But visibility can also be a virtue. Many sources have called attention to the long supply chains of the fossil fuel industries that distance consumers from the sites of production, making the consequences of their consumption largely invisible (see, for example, Mitchell 2011). The pursuit of fossil resources has always been accompanied by all types of violence, from unintentional poisoning to outright murder and war, to propping up dictators, exploitation, theft, cultural genocide, and environmental destruction. The costs of this habit are unevenly distributed to the global south, and to poor and minority communities, while those who enjoy the benefits of energy wealth rarely give a thought to these harsh realities when they flip on the light switch. As political scientist Christopher Jones points out, the seemingly endless supply of energy in what he calls the “mineral energy regime” encourages both ignorance and wastefulness. “The more we use,” he writes, “the less we notice” (Jones 2014: 228). The environmental and health benefits of renewable energy production aside, the high visibility of these technologies would seem to have the additional advantage of making consumers more conscious of their energy use, and the conditions that make their consumption possible.

Many Danes will readily admit their appreciation for these sorts of ancillary benefits, but the democratic process has a way of slowing down development that is unpalatable to ambitious bureaucrats and corporate CEOs. To proponents of agonism, like political theorist William Connolly, the public contestation over proposed wind developments would be viewed as a positive sign of healthy democratic discourse. But Danish authorities, in an effort to avoid these messy and costly disputes, are turning their attention to offshore development. Offshore projects are the sole purview of the national government, bypassing the municipal planning process—and the farther the parks are built offshore, the less likely they are to disturb views from the elites' cherished summer houses. At about 10–15 miles from the coast, wind turbines begin to blend into the horizon, out of sight, out of mind.

Rethinking Revolutionary Technology

Denmark's transition from an energy regime almost entirely dependent on fossil fuels to one that increasingly relies on the wind challenges popular conceptions of the characteristics of renewable energy systems. Wind power has proven flexible enough to be leveraged in all manner of contexts, by all manner of actors, for all manner of purposes. The answer to Langdon Winner's famous question "Do artifacts have politics?" is unequivocally, "yes," though not by virtue of preferences embedded in the technical objects themselves. Rather, artifacts acquire a political orientation by virtue of the larger systems in which they are embedded.

This claim is a familiar one in the philosophy of technology, where the sociotechnical systems theory pioneered by Thomas Hughes has been one of the dominant approaches for the last several decades. A core tenant of this program is that the technical and social components of technology cannot be easily disentangled. In spite of Hughes's influence, and the efforts of constructivists to emphasize the social antecedents of technological change, the aura of the technical artifact still holds an intuitive appeal. There are strategic advantages to targeting the material infrastructure of energy systems—it is easier for sustainability activists to block a pipeline than to bankrupt the oil industry. Proposals for systemic solutions may be too abstract, too uncertain, and too divisive to build the coalitions necessary for action. For analytic purposes, the sturdy materiality of artifacts offers a concrete point of entry. Technical artifacts are far easier to disassemble and explain than the maddeningly complex organizations, discourses, and networks that make up sociotechnical systems—and energy technologies are a paradigmatic case—with properties that cannot be broken down into the sum of their component parts. Artifacts also occupy central locations in energy networks. Wind turbines are direct points of connection to an amorphous grid, and as a result bring many different actors into contact with the larger system. While having an interest in a wind turbine is not a prerequisite for participation in the energy system, nor does it guarantee a substantial role, it does make one a player of some sort. But the importance of artifacts as network hubs may be overstated. They may not be the most revealing units of analysis for understanding how systems function and how they change, and constructing narratives around artifacts may obscure processes occurring at other levels.

Technical questions about the design and performance of wind turbines ceased to be significant constraining factors in development after the 1980s, when now-standard solutions to the major engineering challenges were devised. Since then, technical innovations have largely been restricted to iterative improvements in subcomponents and more efficient organization of construction, maintenance, and manufacturing. The evolution of the wind sector in recent decades has been driven primarily by broader policy agendas and market forces. The emphasis on technological artifacts in the scholarly literature may itself be an artifact of the disproportionate attention given to the earliest stages of development. It is clear that technological systems continue to evolve—and in the case of the Danish wind sector dramatically so—after they have matured and standard technical solutions have been implemented. Origin stories make for extremely compelling narratives, but drawing conclusions based on these early stages of development, as many histories of wind power in Denmark have done, without accounting for how the growth of the system unfolds over time, leads to an incomplete understanding of the processes at work.

A related bias in research on technology is the prominent role often attributed to engineering design. In an effort to deconstruct the appearance of inevitability technical choices acquire in retrospect, these engineering studies show how technological change is managed by interested agents. Historian Walter Vincenti calls these system shapers "the design community" (Vincenti 1994). While there is no need to deny their agency, placing engineers at the centers of these stories makes it more difficult to see the ways in which their understanding of the choices

available to them are constrained by actors, events, and conditions beyond their control. Even if the concept of a design community were extended—to the breaking point of its usefulness—to include any actor who in any way interacts with the system, such an approach would still fail to account for the often critical role of external environmental pressures in shaping technological change. In a subtle bit of physics envy, social theorists often treat technologies as if they were closed systems, and only the internal dynamics mattered. In practice, defining the boundaries of technological systems is a vexing analytical challenge, as they are always intermingled with other systems, and embedded in still larger systems. Thomas Hughes was able to resolve the problem of boundaries only by defining a system in terms of the elements under the control of its managers, which would seem to place many potential explanatory factors outside the system, assuming the question of who counts as a manager can be resolved. Still, Hughes recognized that environmental conditions could exert selective pressures on technological developments.

In real-world energy systems like electric grids, these environmental effects—not to mention nonlinear system effects—are pervasive, but tracing those ripples through technological networks and identifying the forces that produced them will test the mettle of any analyst. The understandable, if not entirely defensible, impulse to focus on the more tractable subjects of artifacts and designers encourages an unwarranted expectation that technical innovation will drive system change. This presumption is an updated version of the old technological determinism, except it is now recognized that the technical systems responsible for shaping society are themselves the products of human agency. This vicious circle ends, at best, with an admission of the co-production of technology and society, which is not so much incorrect as it is a dead end for inquiry.

The philosophy of technology has been grappling with large sociotechnical systems for more than three decades, and its practitioners are still not entirely comfortable with the strange hybrid creatures their research is revealing. Fully appreciating the systemic character of technology, which cannot be reduced to artifacts with social antecedents and components, requires a paradigm shift which may be under way, but remains incomplete. Scholars have developed many creative proposals for describing the ontologies of advanced technologies, but much work remains to be done in improving the commensurability of these theoretical tools and testing the scope of their applications. Bruno Latour has won many converts to his actor-network approach with a series of books and articles challenging modernity's division of nature and culture. While Latourian concepts like "matters of concern" are tantalizingly original, his attempts at grand theory building (which he rather immodestly refers to as a "cosmology") require of readers a substantial suspension of disbelief and prejudice. His distinctive brand of political ecology is among the least literal-minded of numerous calls to study technologies more like ecosystems. Another quasi-ecological perspective can be found in the Multilevel Perspective originating in the Dutch Transition Management literature, which borrows the niche concept to explain radical innovation.

However promising the notion of bridging the gap between ecology and social studies of technology may sound in the abstract, efforts to flesh out precisely how technologies function as ecosystems, in theory and in practice, are still in their infancy. The development of an ecological approach to studying technology remains antithetical to the methods of mainstream social science, which were designed to break the objects of inquiry down into their constituent parts, to isolate variables and trace linear chains of causation. It is not surprising, then, that a theoretical perspective that emphasizes nonlinear and systemic interactions would meet with skepticism. The destabilizing unfamiliarity, the seeming oddness, of such approaches should not recommend their dismissal, but suggest their promise. As the old paradigms become increasingly brittle and ill fitting, philosophers should be unafraid to engage in what Paul Feyerabend describes as "the determined production of nonsense." "Madness," Feyerabend writes,

“turns into sanity provided it is sufficiently rich and sufficiently regular to function as the basis of a new world view” (Feyerabend 1975: 270). Seeking inspiration from the field’s many candidate theories, philosophers of technology must continue to expound a new conceptual language better suited to handling complexity, to emphasizing processes over products, to expecting dynamic change rather than stable equilibrium, and to capturing new expressions of power, agency, and interdependence.

This work will necessitate the deepening of collaboration across traditional disciplinary divides, as understanding multidimensional technological systems will require many diverse types of expertise. It must engage with real-world systems—which is why this discussion has drawn so much on the case of wind power in Denmark—since technologies cannot be isolated from the contexts of their use. The only productive place for a philosophy of technology to begin today is not from first principles or first impressions, but somewhere in the middle, in situ, in process. Researchers of energy systems cannot stand outside the systems they study and claim the position of disinterested observers. The already prominent normative strain in the energy transitions literature must continue making the case for moving beyond dependence on fossil fuels, and evaluating the consequences of pursuing alternative transition pathways.

Whether or not this transition will amount to a technological revolution is first of all a systemic question—it implies a shift more fundamental than merely swapping one technical apparatus for another. But it therefore also requires an ethical stance on the scope of the change necessary or desired.

The Danish experience with wind power demonstrates that, with patience and dedication, the technical impediments to moving toward a renewable energy regime can be overcome, and a significant transformation can be implemented reliably, safely, efficiently, and without the need for individuals to make enormous sacrifices in their lifestyles.

This progressive integration of renewable energy into existing social systems hardly seems revolutionary, but current trends in the Danish energy system are open to a range of interpretations, each of which suggests avenues for future research. If the revolution the Tvind group and its admirers were hoping to instigate has, indeed, failed, then a postmortem investigation could shed light on the causes of its failure. It is highly unlikely that the social vision of wind energy enthusiasts has been totally co-opted, and that wind technology has been absorbed into the existing energy regime without altering it in the process. Documenting and measuring the effects of incorporating wind power can yield insights into how systems adapt to challenges from radical upstarts.

Another possibility is that bigger changes are still to come, and reaching the goal of a 100% renewable energy system in the next few decades will be a whole new order of challenge, requiring more qualitative shifts in Danish society. In this eventuality, assessing the revolutionary character of renewable energy technology at the present stage of development would be premature. Such a trajectory would challenge widely held ideas about the importance of disruptive inventions, tipping points, and technological momentum. The emphasis on innovation studies in recent decades has led to a deficit of knowledge about the latter stages of technological change, and how mature technologies continue to evolve.

One additional possibility is that expectations of technological revolutions may need to be recalibrated. As Arendt has shown, the concept has had widely divergent connotations in different historical epochs, and it may be time for an update. The romantic notion of revolutionary change has been handed down from the scientific revolution, the industrial revolution, and the democratic uprisings of the eighteenth century. It is not only possible, but probable, that the transition to a renewable energy system will not look like any of those revolutions. Denmark’s transition has not been marked by sudden, destabilizing upheaval, but that does not mean the consequences of adopting renewable energy have not been dramatic and far-ranging, and will

not ultimately lead to deep-seated and lasting changes in Danish society. The character of the shift does not fit the model of punctuated equilibrium, but instead appears to be more of a gradual evolution. If that trend continues, it would be consistent with the findings of recent studies on past energy transitions. Christopher Jones has shown that the discovery of fossil fuel resources in the United States and the development of the technical means to exploit them did not lead to their immediate or widespread adoption. Rather, it took several generations for fossil fuels to fully penetrate U.S. society (Jones 2014). While the number of high-quality studies of past transitions continues to multiply, scholars of technology have never had the opportunity to investigate an energy transition in progress, until now. Future research will have to ask what lessons from the past can be applied to the present moment, or to what extent the transition to a renewable energy system really is something new under the sun.

The desire for a novel order, a new beginning, goes hand-in-hand with the notion that revolutions are discrete moments in time; they have clearly demarcated beginnings, and, sooner or later, they end. Revolutionaries are frequently guilty of envisioning utopian outcomes to their struggles, and the environmental movement is no exception. The goal of sustainability is usually understood as some sort of stable end state, a harmonious balance that can be maintained in perpetuity. The appearance of stasis is always an artificial construct, reliant on bracketing the ways real systems—which are never as pure as the categories invented to describe them—are continually evolving. Technologies and the larger systems in which they are embedded constantly shift, creating new constellations of actors, imposing new constraints and opening new possibilities, mutating into new forms with emergent properties.

Revolutionary strategies must be correspondingly flexible and opportunistic. Revolutionaries cannot become too committed to any specific agenda or toolbox, but must be willing to adapt their practices to respond to changing conditions and new obstacles. “Technological revolution,” Larry Hickman wrote, “is not a goal but a process that takes on new dimensions and new import at each stage of its development and whose outcome can therefore never be predicted” (Hickman 1996). Viewed in this way, the success or failure of the revolution the Tvind team hoped to kick start with its turbine project does not depend on how much of the particular social vision it articulated has come to pass since the 1970s. Rather, its efforts were revolutionary in the way it mobilized actors and drew on the resources available to them to find creative solutions to problems that looked insurmountable at the time, and inspired others to carry out their own projects in a similar spirit. Their program ceased to be revolutionary when it stopped functioning as a catalyst for continued technological progress. The model of development that Danish wind enthusiasts pursued in the 1970s may not be replicable today, but that does not mean the revolution is over. The revolutionary spirit is renewable so long as it can serve as an impetus for action on the unique challenges of the present.

References

- Arendt, H. (1963). *On Revolution*. New York: The Viking Press.
- Bostrom, Nick. (2007). Technological Revolutions: Ethics and Policy in the Dark. In N. Cameron and M. Mitchell (eds.) *Nanoscale: Issues and Perspectives for the Nano Century*. Hoboken, NJ: Wiley, 129–152.
- Drucker, P. (1961). The Technological Revolution: Notes on the Relationship of Technology, Science, and Culture. *Technology and Culture*, 2, 342–351.
- Feyerabend, P. (1975). *Against Method*. London: Verso.
- Hickman, L. (1996). Techné and Politeia Revisited: Pragmatic Paths to Technological Revolution. *Techné: Research in Philosophy and Technology*, 1, 116–127.
- Jones, C. (2014). *Routes of Power: Energy and Modern America*. Cambridge, MA: Harvard University Press.

- Klein, N. (2014). *This Changes Everything: Capitalism vs. the Climate*. New York: Simon & Schuster.
- Lovins, A. (1977). *Soft Energy Paths: Toward a Durable Peace*. Washington, DC: Friends of the Earth International.
- Mathews, J. (2015). *Greening of Capitalism: How Asia Is Driving the Next Great Transformation*. Stanford, CA: Stanford University Press.
- Mitchell, T. (2011). *Carbon Democracy: Political Power in the Age of Oil*. London: Verso.
- Raman, S. (2013). Fossilizing Renewable Energies. *Science as Culture*, 22, 172–180.
- Vincenti, W. (1994). The Retractable Airplane Landing Gear and the Northrop ‘Anomaly’: Variation-Selection and the Shaping of Technology. *Technology and Culture*, 35, 1–33.
- Winner, L. (1986). *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: University of Chicago Press.

Geoengineering and Climate Change

Tina Sikka

From the standpoint of a higher socio-economic formation, the private property of particular individuals in the earth will appear just as absurd as the private property of one man in other men [i.e., slavery]. Even an entire society, a nation, or all simultaneously existing societies taken together, are not the owners of the earth. They are simply its possessors, its beneficiaries, and have to bequeath it in an improved state to succeeding generations as *boni patres familias* [good heads of the household].

(Marx 1976, 959)

Geoengineering can be defined as the deliberate technological intervention into our climate in order to limit and/or reverse climate change. It is currently being explored by governments, scientists, universities, and private think tanks as an option in humanity's fight against global warming. In this chapter, I argue against choosing geoengineering as a tactic in dealing with global warming because of (1) substantial flaws in the assumptions that ground economic-centric studies and (2) problems with how data have been selectively used to support its implementation. Many of these studies are erroneously based upon static conceptions of consumption rates, human behavior, and basic social objectives. I also make the case that there is an alternative way to approach the study of climate change and environmental issues using Andrew Feenberg's philosophy of technology approach (Feenberg 1995, 1991/2002, 1999).

Following a brief definition of geoengineering, I begin with a brief discussion of the now indefinitely postponed Stratospheric Particle Injection for Climate Engineering (SPICE) project, which aims to test the feasibility of one specific geoengineering technique. I use this particular case, and a considered analysis of the studies that surround it, throughout this chapter as a way into an overarching analysis of the problems of economic methodologies that are most often applied to the assessment of climate change as well as geoengineering. I begin my examination of these flaws with a brief critique of quantitative methodologies, as they relate to the study of the environment, and follow this with a critical examination of demand estimates using the cases of the airline industry as an example. Next, I discuss the complications associated with the so-called Energy Paradox, which leads to host of further difficulties such as direct and indirect rebound effects. Also reviewed in this section are the problems connected to *ceteris paribus* logic, which assumes that all variables involved in economic analysis of environmental issues will remain unchanged—which they rarely do.

A second set of complications associated with economic studies of the climate change discussed in this piece emerges out of Andrew Feenberg's work on the environment in which he discusses *ceteris paribus* modes of analysis in detail. Feenberg argues that *ceteris paribus* thinking, as it has been applied to the subject of climate in particular, fails to consider the articulation, development, and implementation of new technical standards, which, if widely adopted,

could alter current projections around emissions. While Feenberg's critique shares several assumptions with the economic critique also discussed, it differs in that his approach highlights how changes in technical standards related to the environment might develop over time.

In the third section, I discuss a further flaw in economic studies of climate change that center around the belief that there will be no substantive cultural and/or political constraints on future energy use. The assumption here is that the public will not consent to make changes in lifestyle and consumption patterns in order to mitigate global warming and, consequently, geoengineering will be necessary. This conclusion is not inevitable. What is required, and can develop if given the needed support and space, is a reconfiguration of value hierarchies based on a long-term view of social and environmental well-being.

In the final section of this chapter, a more detailed discussion of how the articulation of new environmental and technical standards and values can engender a transformed relationship with the environment is undertaken. Democratic rationalization, as discussed by Andrew Feenberg in his work on the philosophy of technology, is the end result of this process. Overall, I conclude that the oftentimes painful arguments over technological solutions to environmental problems, and the many possible avenues by which we might begin to deal with anthropocentric climate change, is a natural part of the process of technical evolution.

Moreover, I make the case that the traditional economic arguments that support geoengineering may well come to be widely seen as unsuitable due to the unique nature of climate change. Anthropocentric global warming, in this context, has to be thought of as a cultural issue threatening civilization, not one of economic costs and benefits.

Geoengineering

Geoengineering can be defined as intentional technological intervention into the world's climate in order to prevent, mitigate, or reverse anthropogenic global warming (Allenby 2010; Long et al. 2012). Its techniques are purposeful, have far-reaching global effects, and rely on technology as its central driver. Geoengineering generally falls under one of two categories: solar radiation management (SRM) carbon dioxide removal (CDR). SRM techniques aim to reflect sunlight away from the earth through such methods as enhancing cloud whitening, in which seawater is sprayed into the atmosphere with the aim of increasing the proportion of light-reflecting cloud cover, and albedo enhancement, which involves covering regions with light-reflecting white materials or pumping aerosol sulphates in the stratosphere in pursuit of the same objective. In contrast, CDRs aim to remove carbon dioxide from the atmosphere. In one technique, called *biochar*, geoengineers would plant, burn, and bury large quantities of biomass. This method aims to suck in and store a large amount of atmospheric carbon dioxide during the burning phase.¹ In another CDR method, carbon capture and sequestration, carbon is physically captured and sequestered underground. This technique is currently being tested in the private sector. In addition, ocean fertilization, wherein the ocean is seeded with iron with the objective of increasing the quantity of carbon-absorbing algae blooms, also falls into the category of CDR.

SPICE Project

Currently, a concerted effort is taking place to carefully and comprehensively test the feasibility of SRM geoengineering in the United Kingdom through intensive modeling. The project, called SPICE (Stratospheric Particle Injection for Climate Engineering) is a collaboration among Marshall Aerospace (a European company specializing in military, civilian, and commercial aircraft engineering) and the universities of Cambridge, Oxford, Bristol, and Edinburgh. The project aims to initiate the first steps towards testing the feasibility of injecting light-reflecting particles

into the stratosphere—which will then, it is believed, lead to global temperature cooling. Those involved in the project had, in fact, originally intended to physically test aspects of this geoengineering technique. However, as a result of strong public opposition, the testing elements of the project have been put on hold indefinitely. Yet SPICE remains the most legitimized, well-known, and publicly discussed geoengineering project. As such, it serves as a useful case study—particularly with respect to how it has been justified in economic terms.²

Basically, and consistent with SPICE's SRM geoengineering project, researchers hope that the effects of dispersing sulfur dioxide particulates, or other such aerosols, into the lower atmosphere will reflect sunlight back into space and initiate a cooling effect (note that this would not, however, reduce the existing levels of atmospheric greenhouse gases). Over a number of weeks, the sulfur dioxide would combine with water and oxygen to form sulfuric acid gas and then condense into aerosol droplets, creating a visible haze that reflects sunlight. This method of geoengineering relies on the cooling effects experienced after volcanic eruptions.

The most studied and cited example of such an eruption is that of Mount Pinatubo in 1991. A major study of the Mount Pinatubo eruption in the Philippines was undertaken by NASA's Goddard Institute for Space Studies, which concluded that the net effect of this volcanic eruption was that the released sulfur dioxide, over one year, led to a cooling effect over the Northern Hemisphere of 0.5 to 0.6 degrees Celsius. However, it should also be noted that similar studies show that this was followed by unusually rapid ozone depletion rates in the southern hemisphere (Pinatubo Volcano Observatory Team 1991).

The SPICE project, as stated, marks the preliminary steps towards testing whether the cooling experienced after Mount Pinatubo can be replicated through human intervention—i.e., by injecting sulfur dioxide, or other such particulates, into the atmosphere directly (although this test would have used water). The SPICE test, which was to have taken place in Norfolk, United Kingdom, was to have attached a helium-filled balloon to a 1 kilometer piece of hosepipe that is firmly tethered to the ground. Water would then be pumped from the ground to the top of the hosepipe and sprayed into the surrounding area. According to the lead researcher of the project, Dr. Matt Watson of Bristol University, SPICE, it

will eventually be the first UK project aimed at providing some much-needed, evidence based, knowledge about geoengineering technologies. The project itself is not carrying out geoengineering, just investigating the feasibility of doing so. We hope that by carrying out this research we will start to shed light on some of the uncertainties surrounding this controversial subject, and encourage mature and wide-ranging debate that will help inform any future research and decision making.

(NERC 2011)

Moreover, this test, according to the principal investigators—who still hope to get permission to go ahead in the near future—aims to further two specific strands of inquiry. First, they want to test the particle candidates—that is, to consider “what would be an ‘ideal’ particle to inject into the stratosphere.” In doing so, the researchers “aim to identify a particle with excellent solar radiation scattering properties, and consider what potential impacts might be on climate, weather, ecosystems and human health” (NERC 2011). Second, the scientists involved hope to examine the tethered-balloon transportation/delivery system and build new modeling criteria based on the conclusions reached.

As such, the SPICE project marks the first step towards more focused, small-scale, legitimate geoengineering testing. The next phase anticipates using a 20 kilometer piece of hosepipe, although that particular test is thought to be years away. I return to the SPICE project and methods that have been developed to assess the test's outcomes further on in this piece.

In the next section, a general critique of quantitative research methods is given as they relate to the environment as well as a more intensive and critical assessment of economic-focused, cost benefit modes of analysis—particularly with respect to how they inform environmental issues and geoengineering.

Critique of Economic Approach: Quantitative Methods

More often than not, when studying environmental issues, research methods that are not exclusively quantitative are dismissed as irrelevant because they are perceived as subjective, unscientific, noncumulative, and nongeneralizable. Quantitative ‘scientific’ and empirical data, in contrast, are seen as “unproblematically available to observation free from any prior theoretical commitments” (Demeritt 2006: 455). I, however, am in good company in arguing that there is much more to the study of social, natural, and environmental phenomena than a simple number and discrete sets of data (Van den Bergh and Jeroen 2004; Creswell and Clark 2007). More qualitative modes of analysis must be incorporated into studies of climate change and environmental degradation. The Intergovernmental Panel on Climate Change (IPCC) has taken steps to do so through the incorporation of socioeconomic issues—for example in its 2014 report, which studies issues of food security, poverty, risks, and sustainable development.

Qualitative methods tend to focus more intently on experiences, cultures, understandings and interactions. As Berg argues, “Quality refers to the what, how, when, and where of a thing—its essence and ambience. Qualitative research thus refers to the meanings, concepts, definitions, characteristics, metaphors, symbols, and descriptions of things” (Berg 2007: 3). An overreliance on quantitative methods is problematic in that they are underpinned by a realist, positivist epistemology that assumes there exist objective facts in the world that can be known through specific kinds of scientifically derived and experiment-based research. Overall, using the scientific method (and attendant techniques of observation and deduction) to study the natural and social world is supposed to remove the uncertain and often muddled factors of values, ideology, and politics.

Those skeptical of this perspective, many of whom fall into the category of social constructivists, contend that this is not the case and assert that knowledge about the world is actively constructed. Of particular importance to the subject of geoengineering and climate change is the predictive bias of most quantitative methodologies, which draw on statistical analysis in ways that tend to endow the analyses with an aura of certainty. Quantitative methods also assume a certain predictive and controlling faculty that is far from certain when it comes to the unpredictable side effects of geoengineering, which can range from acid rain to flooding and even drought.

Critique of Economic Studies of Environmental Issues: A General Assessment

Moving from this general critique of quantitative methods to the more specific critique of the kinds of economic analysis associated with environmental impact assessments, it is critical to begin by unpacking the assumptions and biases that permeate analyses of this sort.

First, it is often the case that traditional environmental economics, as they relate to subjects like sustainability and risk, tend to ignore nonmarket values such as the “individual’s willingness to pay for preserving a particular resource for future use, [and] . . . the individuals’ willingness to pay for preserving a particular resource for the sake of its existence” (Venkatachalam 2004: 90; see Straton 2006; Boyd 2007).

Even aspects like the public's willingness to accept environmental risk, which is factored into some more recent cost-benefit studies of environmental policy and technology, reflect a bias towards retaining the environmental status quo when it comes to measures that might upset our current way of life. As such, with respect to the subject of geoengineering, it is often assumed that preserving our current economic and sociopolitical system is of the greatest importance to individuals. Supporters of geoengineering who rely on this assumption often hold an implicit, or even explicit, conservative bias in that they assume a priori resistance to fundamental change. An example of this can be found in the work of Gregory Benford, who, in an essay written for the conservative Reason Foundation in 1997, argues that

Instead of draconian cutbacks in greenhouse-gas emissions, there may very well be fairly simple ways—even easy ones—to fix our dilemma . . . take seriously the concept of “geo-engineering,” of consciously altering atmospheric chemistry and conditions, of mitigating the effects of greenhouse gases rather than simply calling for their reduction or outright prohibition.

(Benford 1997)

This ethos persists today and is reiterated by groups like the American Enterprise Institute (AEI) as well as governments and individual scientists like Kerry Emanuel of MIT and David Keith, formerly at the University of Calgary and now at Harvard.³

Second, it should be noted that much of the disparity between econometric analyses of environmental goods and more socially resonant ones comes down to the way questions about environmental measures are articulated and framed:

Take for instance, a proposed change in an environmental policy that would result in an improved air quality in a particular locality. *Ceteris paribus* (such as property rights, etc.), an individual in the locality can either be asked to state her maximum willingness to pay (compensating variation) for ensuring the change in the policy that aims at say, improving the air quality in the region or she can be asked to state minimum willingness to accept compensation (equivalent variation) required to compensate the expected utility foregone due to nonimplementation of the proposed policy.

(Venkatachalam 2004: 92)

With respect to geoengineering, much of the concern around risk, using economic analysis, has centered on how the calculation of the public's willingness to accept compensation is made and how much compensation would be required to accept the possible fallout. This comes to replace more important questions like the willingness of members of the public to make concrete changes to their lives in order to address the root causes of climate change—rather than relying on a technological quick-fix that may backfire. It also, as Charlesworth and Okereke (2010) argue, undermines the ability for people to make decisions about risk—particularly when they are not given the information needed to do so.

Critique of Specific Economic Methods

In the following sections, I give three further reasons why economic analyses, when applied to the study of the environment and geoengineering, have serious limitations. Such concerns center on problematic demand estimates, the so-called Energy Paradox, as well as an overreliance on flawed *ceteris paribus* logic. The conclusion reached through an examination of these methods

not only points to and supports opponents of geoengineering, but also describes the limits of focusing solely on technological solutions to climate change—even with respect to *specific* energy efficient technologies (even though they are desperately needed)—since these do not address the central problems of an overconsumptive and resource-based economic system.

Demand Estimates

With respect to demand estimates, it is the case that many of the projections used to support the need for geoengineering rely on estimations of energy demands that do not consider possible changes in lifestyle and consumption patterns, or even technological changes made with public input, that might put us on a more sustainable path. These estimates draw on equations that assume the present patterns of emissions, consumption, and demand will continue to rise. It is important to note, however, that statistics related to future energy use are based upon past consumption patterns and are therefore considered by most economists to be probabilities that can be interpreted in a variety of ways. However, I argue that proponents of geoengineering have politicized and transformed these probabilities into projections to support geoengineering.

For example, according to the International Energy Agency's (IEA) research in this area, at present, the following statistics most accurately captures expectations related to future energy use:

Transport accounts for about one quarter of global energy use and energy-related CO₂ emissions. In absence of new policies, transport energy use and related CO₂ emissions are projected to increase by nearly 50% by 2030 and by more than 80% by 2050. Nearly 70% of electricity is generated from fossil fuels: coal (42% of generated power globally in 2007); gas (21%); hydro (16%); nuclear (14%); oil (6%); and non-hydro-renewables (2%). As a result, electricity accounts for 40% of global energy-related CO₂ emissions; these emissions will grow by 58% globally by 2030 unless new policy measures are introduced. Industry accounts for approximately one-third of global final energy use and almost 40% of total energy-related CO₂ emissions. Over recent decades, industrial energy efficiency has improved and CO₂ intensity declined in many sectors, but this progress has been offset by growing industrial production worldwide. Projections of future energy use and emissions show that without decisive action, these trends will continue.

(IEA 2011)

What is striking about these statistics is not how dire they are, but rather that (because they rely entirely on the assumption that consumption patterns will inevitably increase) their assessment of demand estimates has opened the door for some groups to seize on this sense of futility in order to push for drastic forms of technological mitigation like geoengineering.

Some evidence—for example, data collected by the global energy intelligence firm Enerdata—suggests world energy demand is actually decreasing for the first time in thirty years (Enerdata 2010). Note that while the assumption of increasing energy demand and its consequences are not false or faulty, as these figures are often used to support the call for sustainability as well, what is problematic is that proponents of geoengineering have taken to using them to support their case for immediate technological intervention.⁴ By (1) assuming that demand, use, and supply will remain stable (*ceteris paribus*); (2) failing to define the role and drawbacks of extrapolations, which tend to use aggregate data and assume that past trends will predict the future; and (3) by conflating extrapolations with projections, the latter of which assume that certain future conditions will hold based on a set of assumptions about the present, this data then feed into a kind of groupthink on the inevitable need for geoengineering that does little justice to its complexity or tenuousness.

Scientists like Nobel Prize winner Paul Crutzen, Ken Caldeira of Carnegie Mellon, and MIT's David Battisti have all formulated public justifications of geoengineering based on the view that significant reductions in greenhouse gas emissions will not occur and that technologies that will substantively reduce climate change will likely not be conceived of or produced in the near future. In fact, these numbers have functioned as compelling arguments to justify projects like SPICE. Specifically, it has been argued that the SPICE experiment is necessary because of the rapid rate of anthropogenic climate change and a firm belief that consumption patterns will continue to rise unabated. According to the SPICE project's initial grant proposal to the Engineering and Physical Sciences Research Council (EPSRC), it claimed that:

Future projections by climate models indicate substantial changes in future decades, much of which is on a regional scale that will severely impact regions of the world that are already under stress. There has been much improved understanding of the serious nature of the global warming problem both by politicians and the general public in recent years. However, there is great concern that efforts to mitigate future change by reduced greenhouse gas (GHG) emissions, including the outcome of the international meeting in Copenhagen 2009, are proceeding too slowly to avoid the risk of dangerous climate change and the possibility of certain 'tipping points' (such as the collapse of the Indian Monsoon or melting of the Arctic ice sheet) being reached. This has prompted consideration of intervention by alternative means.

(Watson 2010)

Ironically, several private and politically conservative think tanks also hold this view. For instance, in the concluding statement from the Climate Institute's Asilomar on Climate Intervention Technologies, in which many noted scientists took part, it was asserted that: "Despite ongoing efforts to reduce emissions and adapt to the changing climate, global greenhouse gas emissions are far above what is required to reverse the increasing changes in atmospheric composition" (Climate Institute, 2, 2010). Again, while this conclusion is based on sound science, its basic assumption is that nothing behavioral, or even technological (in terms of energy-efficient technologies), will occur to change these figures.

In what follows, I apply the critique of demand estimates to the study of the fuel efficiency in the airline industry. This case offers a concrete illustration of the problems with the conclusions reached using demand estimates based solely on past behavior. Similar conclusions can be extrapolated and applied to the case of geoengineering.

The Airline Industry

An example of in which problematic demand estimates have been relied upon is that of fuel efficiencies related to the airline industry. When evaluated in light of geoengineering, what becomes clear is that (1) energy-efficient technologies need to be assessed and studied using more flexible criteria, and (2) even with the wide implementation of energy-saving technology and regulation (as they exist now), climate warming will likely continue.

With respect to the airline industry, in its 2011 report, the IPCC makes it clear that the current methods used to assess the environmental concerns of the aviation industry (which are then used to guide public policy on efficiency targets and emissions limits) are highly problematic. The report criticizes the overuse of cost-benefit analysis, which tends to conclude that demand for airline travel will inevitably fall as a result of efficiency targets due to higher costs—which are passed on to consumers—thereby harming the airline sector in particular and the global economy in general. This thesis, however, fails to consider the fact that "airlines operate in a

highly competitive environment, and in the short term many may absorb fare increases at a cost to profitability rather than pass them on to passengers” (IPCC 1999). The neglect of factors such as this is rife in economic-focused studies of technologies like geoengineering as well.

These estimates also, the IPCC maintains, do not explain that (1) a good deal of the data in sectors like the airline industry are protected by industry/commercial confidentiality and therefore are not factored into current studies, (2) forecasts require a long appraisal period to adequately calculate the benefits of emissions reductions, (3) scientific uncertainties regarding the impact of emissions leads to the use of estimated pollution reductions as the only measure of benefits, and (4) uncertainties about future trends in technology, which may be much more efficient and safe, are rarely considered (IPCC 1999).

As stated, many of the pitfalls of economic analysis discussed with respect to the airline industry apply to geoengineering studies as well. Some of these problems can be seen in the SPICE project’s public consultation process. While its final report, titled “Public Engagement on Geoengineering Research: Preliminary Report on the SPICE Deliberative Workshops,” offers a nuanced and open discussion of the SPICE project’s objectives, methods, consequences, and the public’s concerns, its basic assumptions ignore the possibility that new, more democratic technologies might be developed or that consumption patterns might fundamentally change. It assumes, as such, that environmental conditions will remain as they are. These shortcomings persist despite the fact that this particular consultative process drew on a qualitative research process, namely the focus group.

Philosopher Andrew Feenberg terms this kind of logic consistent with *ceteris paribus* reasoning, which assumes that all factors considered in these kinds of analysis will remain constant—including demand. While *ceteris paribus* assumptions simplify complex economic data, they also constrain and leave out a large number of important variables and, in doing so, often lead to projections that are divorced from reality. An example of this is the Law of Demand, which asserts that “quantity demanded depends negatively on price *ceteris paribus*.” However, while it is widely acknowledged that this is a “very useful and convenient theory,” it only works as long as the “*ceteris paribus* assumption is not ignored” (Bierens and Swanson 2000). We also have to also consider, as Bierens and Swanson argue, “that complements as well as substitutes exist for most traded goods” (Bierens and Swanson 1998: 4). A wider account and critique of *ceteris paribus* logic follows.

The Energy Paradox

A second misuse of economic logic based on these assumptions leads to what is called the *Energy Paradox*, which is comprised of two interconnected theses.

First is the misperception that it is the consumers themselves who apply unreasonable hurdles to impede investment in energy-saving technologies. This part of the paradox claims that it is the consumer’s inflated monetary expectations, and their frustration when these expectations do not materialize, that lead to objections when the next generation of efficient technologies are made available. This assumption, however, fails to account for the consumer factoring in the high future monetary and social costs of increased energy consumption, which are often not made clear or concrete. For this to be rectified, what would be required is that (1) clear information be made available to consumers about potential costs and savings in the long term; (2) that nonrational behavioral factors and choices be considered, including resource and time constraints as well as the tendency to be “biased toward the status quo,” overall risk aversion, and an increased willingness to “take risks to avoid losses than to achieve gains in making economic decisions” (Pew Center 2011); and (3) that carbon emissions themselves be priced so

that decisions about energy use and efficiency are placed in a social, rather than just a personal or individual, context.

The second and most intractable part of this so-called Energy Paradox asserts that energy-efficient technologies, on the one hand, often create a rebound effect whereby moves towards energy efficiency—like in the case of improved housing insulation as well as energy-efficient cars, appliances, and planes—do not lead to a decrease in consumption but an *increase* due to a decline in the cost of operating the utility (i.e., heating) or appliance. This is known as a “direct rebound effect” and can also include cases in which companies take their profits from investment in energy-efficient technologies and use them to expand the company in ways that not only increase output, but also emissions.

Indirect rebound effects, on the other hand, include cases in which, for example, families who save money on energy-efficient appliances then use the saved money to engage in activities that leave a significant carbon footprint (such as taking a vacation that involves plane travel). It should be noted, however, that this can also occur in circumstances apart from choices made about purchasing energy-efficient technologies when, for example, individuals who save money by buying solar panels use the saved money to purchase something that is emissions-intensive and produces waste, like a newer TV.

Yet, as in most cases in which economic-based analyses are chosen to evaluate complex human behavior, there exists a further often overlooked set of problems related to rebound effects. To begin with, there is competing data on the precise levels of rebound effects, which suggests that there are limits to its applicability to such things as water use, larger appliances, and heating (Greening et al. 2000). Even the less-than-progressive Breakthrough Institute, a think tank that recently compiled a comprehensive literature review of data on rebound effects, argues that the highest levels of rebound occur

not at the consumer level but in the productive sectors of the economy (industry and commerce). Improving the efficiency of a steel plant may result in lower cost of steel, greater demand for steel, and also create greater economic growth—all of which will drive significant rebound in energy use following efficiency improvements.

(The Breakthrough Institute 2011)⁵

The UK Energy Research Centre (UKERC) reached similar findings in a 2007 report titled, “The Rebound Effect: An Assessment of the Evidence for Economy-Wide Energy Savings from Improved Energy Efficiency.” In it, the authors provide a comprehensive assessment of statistics related to perceived rebound effects by comparing a future scenario in which the IEA’s efficiency recommendations were adopted with one in which they were not.⁶ One of the report’s most significant conclusions is that when applied to the household energy use, investment in energy efficient technologies can in fact save a significant amount of energy—particularly when coupled with taxation and regulation. The UKERC report also finds that direct rebound effects are likely to “decline in the future as demand saturates. Improvements in energy efficiency should therefore achieve 70% or more of the reduction in energy consumption using engineering principles” (UKERC, vi).⁷ Factors such as the short time span of studies and the lack of consideration of developing nations’ energy use in this context pose limitations.

Applied to the case of geoengineering, the rebound effect argument has been taken to support the thesis that because increased energy efficiency will not significantly alter levels of greenhouse gas emissions, geoengineering will become necessary. As such, any evidence that suggests that consumption will increase, despite competing data that indicates this does not have to be the case, has to be used to support further testing and research. Physicist David Keith, in a

recent interview with *Scientific America*, makes this very argument. He states that because “the Arctic is melting faster than people expected,” and since not much has been done to counteract this, that geoengineering will have to be adopted (Biello 2011).

However, it is equally important to acknowledge that evidence supporting the thesis that energy-efficient technologies will not solve the climate problem can be interpreted in another way: i.e., to suggest that *existing* energy efficient technologies can only do so much, that more open and publically initiated discussion of new technologies is required, and that fundamental political and economic change is necessary. Accompanying this line of thought, a case can also be made for the argument that an overreliance on technological solutions like geoengineering will likely fail to solve the underlying problem: our extractive and consumption-based economic system.

Ceteris paribus

The third and final economic argument that has been used to undermine less-invasive technological solutions to climate change is an overreliance on *ceteris paribus* logic. It is significant, with respect to the *ceteris paribus* projections and the supportive conclusions derived thereof, that this mode of reasoning has been consistently drawn on to support both SRM and CDR geoengineering. This is particularly evident with respect to discussions taking place at the levels of governance and public policy such at the Royal Society (2009) and the U.S. House Committee on Science and Technology (2010). Both of these bodies have reached conclusions that support the *possible* need for geoengineering based on statistics relating to future energy use and have agreed to work together on geoengineering issues.

Since 2009, both bodies have undertaken research on international decision making, governance of geoengineering research and deployment, and on the science, price, and environmental impacts of these technologies. While research in this area continues, even when based upon the most lukewarm support for further study, it is most often based upon the assumption that everything will remain *ceteris paribus*. Geoengineering’s advocates have then taken this up not as evidence of the need for immediate changes to rates of energy consumption and economic formations, as many in the scientific community contend, but to support geoengineering as a viable alternative to more traditional mitigation strategies. It recently came to light that Russia asked the IPCC to include a discussion of geoengineering in its 2013 report—particularly in light of the dire *ceteris paribus*-based projections coming out of the IPCC. Russia has also publicly acknowledged that it is conducting its own geoengineering research.

The 2013 IPCC report itself, in a box marked TS.7, provides an overview of geoengineering and highlights the unknowns of both SRM and CDR methods. While both critical and skeptical of both approaches, the report also suggests that, *ceteris paribus*, “Modelling indicates that SRM methods, if realizable, have the potential to substantially offset global temperature rise” (IPCC 2013, 21). Critics have highlighted that the inclusion of TS.7 serves to legitimate geoengineering and that this tepid statement, which is followed by several caveats, gives backing to its proponents.

In closing this section, it is important to note that the use of quantitative methods to assess environmental issues and technologies, like climate change and geoengineering, also tends to influence ethical analyses. Most ethical examination of geoengineering draws on a consequentialist ethical framework that reduces normative considerations to how economic effects are distributed (Jameson and Elliot 2009). In this way, even the atmosphere comes to be seen as property (i.e., a scarce resource) subject to our use. This kind of instrumental logic relies on a drive for human mastery over nature which, philosophers of technology like Andrew Feenberg argue, negates human reason and autonomy, and endangers human survival.

Feenberg's Philosophy of Technology Approach

Feenberg's approach to environmental issues is rooted in a critical philosophy of technology and builds on these arguments. He makes the case that calculations of efficiency and cost-benefit analysis, with respect to the adoption of technologies, fail to account for the complex social and cultural forces that guide technological change. What makes Feenberg's approach distinct from other critiques of economic-centric studies of climate and climate technologies is that it is rooted in a conception of technical rationality and founded on the need for further democratization.

Briefly, Feenberg views technology as an inherently social artifact that is replete with meanings that are not entirely "inscribed into the nature of the technology" itself (Feenberg 2010: 15). As such, he contends that technical choices can be made that reflect socially resonant and ecologically sensitive cultural values and meanings rather than standards of abstract control, efficiency, and economic power. This was the case with the internet, which—rather than serving as a mere portal to access to information or for the purposes of emergency communication as was intended—became a nexus of social networks and interaction. In this context, democratic rationalization, Feenberg argues, can be understood as the process by which the controversies that surround technology are funneled into the production of innovative solutions that do not reflect power, money or abstract rationality but "a struggle to subvert the technical practices, procedure and designs structuring everyday life" (Feenberg 2010: 27).

In line with this approach, Feenberg, in a piece titled "Incommensurable Paradigms: Values and the Environment," from his 2010 book *Between Reason and Experience*, draws on this critical framework to claim that we must reject measures of technical rationality that are defined in purely economic terms since these kinds of calculations hold economic boundaries to be invariant. Historically, however, it is clear that precisely what should be subject to economic assessment are the changing nature of these very boundaries. In this sense, it could be argued that the current controversy surrounding geoengineering is in fact part of the natural evolution of technical norms and boundaries which could, at some point, place environmental issues in the realm of social, rather than economic, interests. I return to this point further on.

One of Feenberg's central arguments against the sole use of economic analyses of environmental issues lies in this conception of boundaries, wherein the environment is forcibly placed within the boundary/sphere/purview of the economy where it does not necessarily fit and where solutions like geoengineering become common sense or, as Feenberg puts it, certain technical solutions "rigidifies into destiny" (Feenberg 1999: 14). Feenberg also, in doing so, concludes that it is impossible to

place monetary value on such things as natural beauty and good health because, but these values have been translated into monetary terms to enter the calculation. Trade-off arguments are thus often based on flimsy estimates of costs and benefits when they are not ideological expressions of hidden interests.

(Feenberg 2010: 33–34)

This applies directly to the case of geoengineering as well.

Moreover, according to this constructivist approach, which incorporates a clear account of democracy and participation, environmental values can and should be "incorporated into technically disciplines and codes" such that environmentalism is accepted as "a self evident advance" (Feenberg 2010: 43). This conception of democratic rationalization and its application to environmentalism and technology is critical in that it provides a clear standard by which to assess geoengineering and formulate new goals. Economically based models of analysis, as

demonstrated, do not allow for this because the economy has a boundary and cannot account for the social values needed to think about the use of technologies like geoengineering or even issues around food safety, rainforest protection, or nuclear energy. Moreover, this theory of rationalization, particularly with respect to the formulation of new objectives and knowledge—and when coupled with Feenberg’s considered rejection of cost-benefit analysis—offers a corrective to traditional way new technologies have been studied.

Turning back to geoengineering, it is significant that the majority of studies, commentaries, and justifications of its further research and potential use rely on precisely the kinds of economic assessments Feenberg criticizes. A significant example of this is the work of the Copenhagen Consensus Center (CCC)—a policy think tank funded by the Danish Ministry of Foreign Affairs—that aims at developing studies of the world’s major challenges using the expertise of preeminent economists. All of the group’s major studies employ traditional and rather conservative economic approaches. In its most recent study of climate change mitigation, twenty-four papers were commissioned from climate economists to answer the following question:

If the global community wants to spend up to, say \$250 billion per year over the next 10 years to diminish the adverse effects of climate changes, and to do most good for the world, which solutions would yield the greatest net benefits?—i.e., what are the costs and benefits of different viable climate interventions . . . given some reasonable assumptions about sensible policies for the rest of 21st century?

(CCC 2009)

In a widely cited and highly ranked paper submitted to the CCC for this project, under the aegis of the AEI, J. E. Bickel and Lee Lane make the case for geoengineering in line with the CCC’s call for a purely economic, cost-benefit approach. Their conclusion is as follows:

We estimate that the benefit of a single watt of SRM is worth over \$6 trillion under an emissions control regime of optimal abatement. Furthermore, we show that a single watt of SRM has the same economic benefit as capturing and sequestering almost 65% of yearly CO₂ emissions, which, in conjunction with AC’s significant costs, argues in favor of SRM in the near term.

(Bickel and Lane 2009: 3)

Most of Bickel and Lane’s study stresses the significance of quantitative benefits, cost estimates, transaction costs, and political market failures. For example, their list of potential ‘costs’ of deploying climate engineering technologies include the costs of resources to develop, deploy, and fund the technology, potential costs of conflict, and monitoring costs. Even their study of politics remains focused on potential transaction costs—particularly with respect to the potential for “political structures and rules [to] sometimes block or distort the choice of the best response to a problem” (Bickel and Lane 2009: 26). They suggest that stratospheric aerosol injections, of the SPICE project kind, have a very attractive benefit-cost ratio, which should warrant further testing.

This mode of analysis, according to Feenberg’s framework, poses a significant boundary problem since it ought to be the case that the logic of trade-offs, with respect to the climate, be considered irrelevant because of the unique status of the environment. Not only are decisions regarding the natural environment intrinsically connected to moral questions about the kind of world we want to live in, and thus not reducible to economic calculations, but they also cannot be reduced to *ceteris paribus* logic.

Ceteris paribus logic, as Feenberg makes clear, draws on technological determinism and a neutral conception of technology that have “long since been superseded by more sophisticated approaches” (Feenberg 2010: 34). When applied to the geoengineering debate, technological determinism and related assumptions about the neutrality of technology serve to minimize the socially constructed nature of these technologies, whose design and development are a result of social choices that could have gone a number of ways. Also ignored is geoengineering’s tendency towards incorporating techno-centric, control-oriented, and instrumental technical values at its core.

Therefore, by confining the economic measure of trade-offs when studying technology, a more sophisticated notion of historical change can emerge (i.e., one that is not static and therefore resistant to purely economic analyses). This is in addition to the realization that boundaries between what is subject to and what lies outside economic analyses shift with time and, as such, some things, like the environment, should not be subject solely to economic calculations. Yet many of the potential side effects of geoengineering still continue to be quantified primarily in economic terms by scientific bodies, think tanks, and governments.

For example, in the Royal Society’s 2009 report on geoengineering, titled “Geoengineering the Climate: Science Governance and Uncertainty,” the Society draws on the four primary categories of effectiveness, affordability, timeliness, and safety to evaluate geoengineering technologies. Its method of reaching conclusions with respect to the possible deployment of these techniques is to assign the parameters of very high, high, low, and very low to these four categories. The Society describes these respective groupings as relatively static “technological criteria” that are distinct from aspects like “public attitudes, social acceptability, political feasibility and legality which may change over time” (Royal Society 2009: 7). What the Society fails to note, however, is that technological norms and decisions are also subject to change and often these decisions have unintended consequences.

It is noteworthy that the criteria employed to evaluate geoengineering in this report, specifically with respect to so-called technological factors, places each of these categories on a relatively equal footing. As such, even safety—which could be interpreted in relation to the social good and collective risk and thus placed above other factors—is played off against the criteria of cost and effectiveness. In this sense, if a geoengineering technique is seen to be highly effective and low in cost with medium risk to safety, its use could be justified.

This kind of reasoning is deployed in other areas of the Royal Society’s report as well. For instance, according to the study, ocean fertilization techniques are deemed to be acceptable because, while they receive a safety ranking of very low, they rank high in other categories. Yet this fertilization approach is not only unlikely to be effective in removing carbon dioxide from the atmosphere over the long term, but could also lead to an increase in so-called oceanic dead zones, increased acidification of the deep ocean, and conflicts over food security if the phosphorus used for such purposes rises in cost (phosphorus is required for agricultural production).

Cloud seeding is another approach that, while potentially effective in reducing carbon dioxide levels, could affect weather patterns in countries already suffering from unpredictable changes (e.g., by leading to increased precipitation in India and Bangladesh). According to the Royal Society’s own report, stratospheric seeding could negatively affect biological productivity and have adverse effects on the hydrological cycle, on the ozone, and on high-altitude cloud production.

Yet despite these troubling consequences, overall, the Royal Society reaches surprising conclusions about geoengineering in general and CDR techniques in particular. In fact, it goes as far as to argue that some CDR techniques are “safe, effective, sustainable and affordable” (Royal Society 2009: xi) and should therefore be studied further. It is troubling that this monetized

trade-off approach forms the basis of most assessments of geoengineering since it fails, as much of this report does, to reflect the fact that “effects on social wealth [that are] . . . significant to policy . . . must be measured with respect to fulfillment of actual desires, not theoretical constructions” (Feenberg 2010: 44) like affordability, timeliness, safety, etc. It is also worth noting that these approaches are normative and deeply undemocratic as they seek to impose a specific kind of ethical framework not shared by those likely to be affected.

As described in the next section, by drawing on the case of steamboat boilers and child labor, Feenberg illustrates ability of a critical philosophy of technology to reveal the absurdity of limiting our judgment of regulation and technology to static and deterministic economics, rather than environmental, cultural and political criteria.

New Social and Technical Criteria

During the early nineteenth century in the United States, ticket sales for travel on steamboats consistently rose despite a rising death toll, which resulted from problems with dangerous boilers. According to Feenberg, if a purely cost-benefit analysis is employed to this case, there would be no need to increase regulation on boilers since it would appear that passengers had consented to take on the risk in exchange for low ticket prices—which was an argument widely held in debates over regulation at the federal level. However, in the end, the U.S. government unilaterally decided to regulate boilers and, in doing so, “prioritized the prevention of accidents” (Feenberg 2010: 41) over economic factors such as profit.

Feenberg, makes precisely this point in relation to child labor as well, the regulation of which, early on, was objected to because of its potential economic cost. That is, the case was made that the abolition of child labor would have “catastrophic economic consequences—increased poverty, unemployment, loss of international competitiveness—from the substitution of more costly adult labor” as well as from “the depreciation cycle of machinery,” which would have to be replaced in order to accommodate adult workers, and from “lower wages and trade problems” (Feenberg, 2010, 12).

Yet when child labor was abolished, none of these nightmare scenarios came to fruition and, as a result, child labor became seen as incompatible with the values of society. In addition, the concerns about changes to industrial machinery needed for this transformation in labor practices also came to be understood as both necessary and just. Understood in this way, it becomes clear that the social redefinition of technical values and standards, as well as of human progress and development, is possible.

Therefore, when applied to geoengineering and climate change, as in the previous two cases, what becomes essential is need for active support of the formulation of new technical standards and human values that are not based on personal consumption, but on the welfare and health of future generations. Thus at some point, like in the case of child labor, new, more just environmental standards could be articulated, accepted, and normalized in such a way that they no are longer contested and, as a result, take on the air of commonsense—much like it now appears ridiculous and callous to suggest that arguments in support of child labor based on its economic benefit could carry any weight.

Overall, Feenberg, using his philosophy of technology perspective, draws on these examples to demonstrate how social goals are nested in competing and changing hierarchies that are open to interpretation. Following this argument, it is clear that such social goals as national unity, collective safety, and the social good must be considered in decisions related to geoengineering as well. Yet it remains the case that an overwhelming number of contemporary studies on geoengineering and climate change tend to ignore this necessity. For instance, it is significant that the

overriding assumption of the Royal Society's report on geoengineering is founded on an overt rejection of the possibility that energy intensive technologies themselves could, through a concerted effort, be made redundant in the near future. While the report's authors do emphasize the need to increase efforts aimed at reducing greenhouse gas emissions in line with the UN Framework Convention on Climate Change (UNFCCC) criteria (at least 50% of 1990 levels by 2050), the report is clear that present efforts at reaching this goal are failing and, therefore, because

there is no credible emissions scenario under which global mean temperature would peak and then start to decline by 2100. . . . Unless future efforts to reduce greenhouse gas emissions are much more successful than they have been so far, additional action [geoengineering] may be required should it become necessary to cool the Earth this century.

(Royal Society 2009: ix)

What is even more perplexing is that even those opposed to geoengineering tend to use economic logic in their rejection of these radical technologies. For example, in recent meeting of the Ecological Society of America, consensus was reached amongst prominent scientists (both natural and social) that the risks of geoengineering far outweigh its benefits. Robert Jackson, key organizer of the meeting and the director of Duke University's Center on Global Change, makes the case that there are too many unknown factors for geoengineering to be considered a viable option, particularly since "The bigger the scale of the approach, the riskier it is for the environment" (ESA Press Release 2009). Very little is said about the articulation of new values and, by extension, new technical codes.

These flaws with economic-centric arguments make it even easier to avoid the difficult decisions required to change our present resource-intensive and environmentally exploitative patterns of production and consumption. Feenberg asserts that much of the lack of social and institutional drive to make changes are a result of the fact that "our civilization" such that it is, "was built by people indifferent to the environment" and, therefore, "Environmental considerations were not included in earlier technical disciplines and codes" (Feenberg 2010: 43). This indifference to environmental values, when added to the levels of comfort many of us enjoy in the West, not only explains why there is a kind of inertia in making changes that might upset this comfort, but also why, as Feenberg explains, it is so difficult to impose environmental regulations on industry. Even the Royal Society, as noted, focused its evaluation of various geoengineering techniques on the vectors of effectiveness, affordability, timeliness, and safety—giving equal weight to each.

However, Feenberg also makes it clear that while, at present, environmental values are seen as extraneous to and even alien to the current norms guiding technological innovations like geoengineering, it remains possible that, as in the cases of steamboat boilers and child labor, such values as respect for nature and care for the environment could be seen "as a self-evident advance" (Feenberg 2010: 43) in the near future. What is required, therefore, is that the new crop of energy-efficient technologies be founded on technical codes that address the socially and politically grounded problems that gave rise to climate change, while also addressing the values upon which our society is organized—which is to say that if energy-efficient technologies are still technocratically inclined, control-oriented (Franklin 1992), authoritarian (Mumford 1964), and rationalized by traditional economic analysis, very little will change. This is particularly the case since these instrumental values have solidified into technologies that are inflexible and undemocratic. While it is the case that central control is sometimes necessary—as in the case of water, sewage, and energy systems—it is the ability to integrate public input into technical decision making in a democratic way that is sorely lacking.

Conclusion

Overall, what is needed in our struggle against anthropogenic climate change is a firm rejection of technologies like geoengineering that perpetuate rationalizations and technical codes that are instrumentally oriented and, therefore, unable to address pressing ecological, political, and sociocultural needs. Supporting this tendency is a preponderance of contemporary economic analyses of climate change and geoengineering that rely on the politicization of demand estimates, *ceteris paribus* logic, and cost-benefit analysis, which are then used to reach conclusions in support of risky technologies. These methods, I argue, must also change since they not only fail to reflect the changing nature of social life, but also supply a kind of permission to avoid making difficult choices.

In response, as Feenberg and others who subscribe to the philosophy of technology approach argue, what is needed is the articulation of new technical standards and codes based on values that will usher in a technical paradigm that reflects democratic choice. Geoengineering, because its very telos is based on technocratic and instrumental codes, does not fall into this category. This complex, yet far more difficult, tactic would require us to instigate an entire rethinking of our approach to the natural environment, technological growth, and development. New democratic technical codes, according to Feenberg, will then emerge through a process of technical evolution via considered debate, discourse, protests, and legal challenges that have accompanied changes to technology and policy in the past.

To conclude, it must be noted while this transformation will definitely entail the use of renewable energy and new technologies, it cannot be limited solely to this. James Lovelock makes this clear in a 2010 interview with the BBC in which he states that while renewable energy might make good business sense, it remains to be proven whether they can solve global warming (Hickman 2010). Bill McKibbin makes a similar case with respect to renewable technologies and calls for a hefty tax to be placed on carbon emissions:

One of the great side effects of moving to renewable power is that we will replace vulnerable, brittle centralized systems that are too big to fail with spread out democratic energy sources small enough to be resilient. As such, it is clear that significant social, political and behavioral changes are required to address global warming. This, when coupled with proposals like the democratization of the energy grid, which would involve decentralizing its control, could very well form the basis of a comprehensive solution to anthropogenic climate change.

(Battaglini et al. 2009)

Notes

- 1 Biochar is produced when various forms of plant matter are heated in a low-oxygen environment. When the remaining matter (biochar) is placed in soil, it removes carbon from the atmosphere and stores it, thus acting as form of sequestration.
- 2 SPICE studies have also formed the basis of David Keith and James Anderson's (both Harvard engineers) current plan to conduct a similar field experiment in New Mexico.
- 3 Note that Keith is more reluctant than AEI or Emanuel in his support of geoengineering, preferring to pose it as a unfortunate but necessary alternative that must be studied.
- 4 Nor, in this context, does it matter whether these projections are true or false. What matters is how these estimates are taken up without proper explanation and contextualization.
- 5 Many scientists (Evan Mills at Berkeley's Lawrence Berkeley National Laboratory and Jim Sweeny of Stanford) have come out against the Breakthrough Institute's assertions with respect to the rebound effects of corporations, which they argue, are based on faulty research.

- 6 The Research Councils UK, which has a rich history of funding research, funds the UKERC through public monies and block grants (the councils also fund other research groups inside and outside the scientific, technological, and environmental domain).
- 7 This does not, however, account for indirect rebound effects, which might limit this reduction.

References

- Allenby, B. (2010). Climate Change Negotiations and Geoengineering: Is This Really the Best We Can Do? *Environmental Quality Management*, 20(2), 1–16.
- Battaglini, A., Lilliestam, J., Haas, A. and Patt, A. (2009). Development of SuperSmart Grids for a More Efficient Utilisation of Electricity From Renewable Sources. *Journal of Cleaner Production*, 17(10), 911–918.
- Berg, B. (2007). *Qualitative Research Methods for the Social Sciences*, 6th edition. Boston, MA: Pearson Education.
- Benford, G. (1997). *Climate Controls*. Reason Foundation. Available at: <http://reason.com/archives/1997/11/01/climate-controls> [Accessed October 1 2011]
- Beyond Zero Emissions. (2011). Australian Sustainable Energy: Zero Carbon Australia Stationary Energy Plan Synopsis. *Australian Sustainable Energy*, Web. Available at: www.energy.unimelb.edu.au/uploads/ZCA2020_Stationary_Energy_Synopsis_v1.pdf. [Accessed October 10 2011].
- Beillo, D. Saving Nature by Ending It: Geoengineering and the Moral Case for Conservation. *Scientific America*, June 21. Web. October 14, 2011. Available at: <http://blogs.scientificamerican.com/observations/2011/06/21/saving-nature-by-ending-it-geoengineering-and-the-moral-case-for-conservation-video/>
- Bickel, E. J., and Lane, L. (2009). An Analysis of Climate Engineering as a Response to Climate Change. *Copenhagen Consensus on Climate*.
- Biello, D. (2011). The Future of Climate Change. *Scientific America*. Available at: www.scientificamerican.com/report/climate-change-durban/.
- Bierens, H. J., and Swanson, N. R. (2000). The econometric consequences of the ceteris paribus condition in economic theory. *Journal of Econometrics* 95(2), 223–53.
- Bierens, H. J., and Swanson, N. R. (1998). *The Econometric Consequences of the Ceteris Paribus Condition in Economic Theory*. Available at: <http://econ.la.psu.edu/~hbierens/CETPAR.PDF> [Accessed September 15, 2011].
- Boyd, J. (2007). Nonmarket Benefits of Nature: What Should Be Counted in Green GDP? *Ecological Economics*, 61(4), 716–723.
- CCC. 2009. Fix the Climate: Advice for Policymakers. *Copenhagen Consensus Center*. Available at: www.copenhagenconsensus.com/publication/fix-climate-advice-policymakers.
- Charlesworth, M., and Okereke, C. (2010). Policy Responses to Rapid Climate Change; an Epistemological Critique of Dominant Approaches. *Global Environmental Change*, 20, 121–129.
- Climate Institute. Asilomar Conference on Climate Intervention Technologies. *Washington*, November 2010.
- Creswell, J. W., and Clark, V. L. P. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.
- Demeritt, D. (2006). Science Studies, Climate Change and the Prospects for Constructivist Critique. *Economy and Society*, 35(3), 453–479.
- Enerdata. (2010). *World Energy Demand Down for the First Time in 30 Years*. Web. Available at: www.enerdata.net/enerdatauk/press-and-publication/publications/decrease-worldenergy-demand-first-time-30-years.php [Accessed June 2010].
- ESA. (2011). ESA Press Release. *ESA*. Web. September 28, 2011. Available at: www.esa.org/pao/newsroom/pressReleases2009/08062009.php.
- Feenberg, A. (1995). *Alternative Modernity*. Berkeley, CA: University of California Press.
- Feenberg, A. (1991/2002). *Critical Theory of Technology*. Oxford: Oxford University Press, 1991. Repub. as *Transforming Technology: A Critical Theory Revisited*. Oxford: Oxford University Press, 2002.

- Feenberg, A. (1999). *Questioning Technology*. London: Routledge.
- Feenberg, A. (2010). *Between Reason and Experience: Essays in Technology and Modernity*. Cambridge, MA: MIT Press.
- Franklin, U. (1992). *The Real World of Technology*. (CBC Massey lecture Series). Concord: Ontario: House of Anansi.
- Greening, L. A., Greene, D., and Difiglio, C. (2000). Energy Efficiency and Consumption—The Rebound Effect—A Survey. *Energy Policy*, 28, 389–401.
- Hickman, L. (2010). “James Lovelock: Humans are too stupid to prevent climate change.” *The Guardian*, March 29, 2010.
- IEA. (2011). *Fast Facts*. International Energy Agency. Web. Available at: www.iea.org/journalists/fast_facts.asp [Accessed September 12, 2011].
- IPCC. (1999). *Aviation and the Global Atmosphere: Framework for Evaluation*. Intergovernmental Panel on Climate Change. Web. Available at: www.ipcc.ch/ipccreports/sres/aviation/152.htm [Accessed September 15, 2011]
- IPCC. (2013). *Summary for Policymakers*. Web. Available at: www.ipcc.ch/report/ar5/wg1/#.UmgnxL9JXCQ [Accessed September 2013]
- Jameson, D., and Elliot, R. (2009). Progressive Consequentialism. *Philosophical Perspectives*, 23(1), 241–251.
- Long, J. C. S., Hamburg, S., and Shepherd, J. (2012). Climate: More Ways to Govern Geoengineering. *Nature*, 486(7403), 323–323.
- Marx, K. (1976). *Capital*, vol. 1. London: Penguin.
- Mumford, L. (1964). Authoritarian and Democratic Technics. *Technology and Culture*, 5(1), 1–8.
- National Environmental Research Council (NERC). (2011). *The SPICE Project: A Geoengineering Feasibility Study*. NERC. Web. September 28, 2011. Available at: www.nerc.ac.uk/press/releases/2011/22-spice.asp
- Pew Center for Global Climate Change. (2011). *Residential End-Use Efficiency*. Pew Center. Web. October 2, 2011. Available at: www.pewclimate.org/docUploads/Residential%20End-Use%20Efficiency%20final_0.pdf
- Pinatubo Volcano Observatory Team. (1991). Lessons From a Major Eruption: Mt. Pinatubo, Philippines. *Transactions*, 72, 545–555.
- Popper, K. (1963). *The Logic of Scientific Discovery*. London: Hutchinson.
- Straton, A. (2006). “A Complex Systems Approach to the Value of Ecological Resources.” *Ecological Economics*, 56(3), 402–411.
- The Breakthrough Institute. (2011). *Energy Emergence: Rebound and Backfire as Emergent Phenomena*. The Breakthrough Institute, February. Web. Available at: http://thebreakthrough.org/blog/2011/02/new_report_how_efficiency_can-print.html
- The Royal Society. (2009). *Geoengineering the Climate: Science, Governance and Uncertainty*. Royal Society. Web. Available at: <http://royalsociety.org/policy/publications/2009/geoengineering-climate/> [Accessed September 1, 2009].
- U.S. House of Representatives. (2010). *Hearing: Geoengineering*. Committee on Science and Technology. Web. Available at: www.gpo.gov/fdsys/pkg/CHRG-111hhr53007/pdf/CHRG-111hhr53007.pdf
- Van den Bergh, J. C. J. M. (2004). Optimal Climate Policy Is a Utopia: From Quantitative to Qualitative Cost-Benefit Analysis. *Ecological Economics*, 48(4), 385–393.
- Venkatachalam, L. (2004). The Contingent Valuation Method: A Review. *Environmental Impact Assessment Review*, 24(1), 89–124.
- Watson, Dr., M. (2010). SPICE: Stratospheric Particle Injection for Climate Engineering. EPSRC. Web. Available at: <http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/I01473X/1> [Accessed October 12, 2011].

Fracking

Adam Briggles

Conceptual and Historical Overview

In 2014, Merriam-Webster added the word *fracking* to its dictionary. The definition chosen was “the injection of fluid into shale beds at high pressure in order to free up petroleum resources (such as oil or natural gas).” Notably, fracking was one of the few new words attached to the relatively old technology of oil and gas extraction. Other entries signified the rapid cultural and material innovation surrounding much newer technologies, especially ICTs—for example, “selfie,” “hashtag,” “facebookian,” “crowdfunding,” and “big data.” The presence of fracking among this crowd is a reminder that our increasingly dematerialized and ethereal world is still tethered to the earth and dependent on our extractive relationship with it.

With its focus on “the injection of fluid,” the Merriam-Webster definition is narrowly construed. In this sense, fracking is understood as shorthand for *hydraulic fracturing*, a well-stimulation technique that uses pressurized liquids to induce cracks in rock formations in order to maximize recovery of hydrocarbons. The U.S. Environmental Protection Agency has defined this process slightly more broadly to include the “acquisition of source water, well construction . . . and waste disposal” in addition to well stimulation (EPA 2010).

Yet *fracking* is often used even more broadly to refer to the convergence of hydraulic fracturing with other exploration and production innovations like three-dimensional seismic imaging and horizontal drilling. In this way, fracking denotes the entire upstream process of exploration, site development, drilling, and well stimulation. Thus, one can speak of the roughly three-acre area devoted to extraction and production as a *frack site*. In its widest sense, the term can even refer to midstream and downstream activities of transporting, refining, and combusting oil and gas. People speak, for example, of a pipeline carrying “fracked oil” or a power plant combusting “fracked gas” to produce electricity.

The definition chosen for fracking will determine the answer to how long it has existed. If fracking means breaking up oil- and gas-bearing rocks, then it has been around since the 1860s, when drillers in Pennsylvania used explosive nitroglycerin to increase oil production. A century later, in 1967, the U.S. Atomic Energy Commission attempted the most powerful version of explosion-based fracking by detonating a 29-kiloton nuclear bomb 4,200 feet below the surface of northern New Mexico. This Project Gasbuggy did increase production from the Lewis Shale formation, but the gas was too radioactive to be useful (Carlisle and Carlisle 1967).

The use of high-pressure fluid-based fracking (thus nonexplosive, although perforation guns are still required) began in the 1930s. Engineers experimented with several techniques and fluids, including acids, thickened gasoline, diesel, gels, water, and foams. In 1947, Stanolind Oil and Gas Company conducted the first experimental treatment of a nonacid hydraulic fracturing (called *Hydrafrac*) in Kansas, using 1,000 gallons of a thickened gasoline. That experiment was largely a failure. But by 1950, Halliburton Oil Well Cementing Company had hydraulically fractured 332 wells, with an average production increase of 75% over conventional, nonfracked

wells (Montgomery and Smith 2010). It used a mixture of crude oil and gasoline (averaging 750 gallons of fluid per well) and sand (averaging 400 pounds per well) to prop open the fractures induced in the rocks.

As impressive as this progress was, fracking was not yet powerful enough to produce commercial quantities of oil and gas from unconventional reservoirs. Conventional reservoirs of oil and gas are relatively loosely bound, such that they tend to flow with little or no stimulation. Those reservoirs originated from ‘source rocks’ like the Barnett, Bakken, Eagle Ford, Niobrara, or Marcellus shale formations. Such tight formations are often called the *kitchen*, because that’s where the oil and gas are originally ‘cooked.’ The hydrocarbons in shale formations are locked in extremely tight pores within the rocks, making them very difficult to extract.

By the late 1970s, U.S. oil and gas production appeared to have peaked, and President Jimmy Carter gave a nationally televised address informing Americans that the nation was “simply running out” of reserves. Of course, that was only because the technology did not yet exist to harness the unconventional reserves locked in shale formations. Thus began an enormous R&D effort in the United States jointly conducted by the private and public sectors (see DOE 2007; DOE 2009; Wang and Krupnick 2013). Energy research on fossil fuels grew by an order of magnitude between 1974 and 1979. The goal was to figure out how to extract oil and gas from unconventional reserves.

There were major technological hurdles to overcome. For example, increasing surface area contact with the shale formation requires the ability to turn the wellbore from its vertical entry into the earth into a horizontal position. Among many other innovations, this required the development of sophisticated computing and imaging (see Yost and Overbey 1989; EIA 1993).

Inventing a sufficiently powerful suite of fracturing techniques was also a daunting challenge (see Steward 2007). Throughout the 1980s and 1990s, engineers added more chemicals (now roughly 750 in total) to the frack fluid to serve a variety of purposes (manage viscosity and pH, kill bacteria, prevent corrosion, etc.). Eventually, water became the fluid of choice, which, under very high pressures (roughly 4,500 psi applied by an array of engines on pumping trucks), is able to create the desired fractures provided enough of it is used (on average, a fracked well now requires roughly 4 million gallons of water). More sand (with many wells requiring over 4 million pounds) was also used to prop open the induced cracks in order to enable gas and oil flow.

By 1998, Mitchell Energy and Development Corporation in north Texas had refined a hydraulic fracturing method that was capable of yielding commercial quantities of natural gas from the Barnett Shale. A few years later, Devon Energy (which had acquired Mitchell) married this hydraulic fracturing method to the latest advances in horizontal drilling. By the turn of the twenty-first century, oil and gas operators had ‘cracked the shale code,’ and a major boom in energy production ensued, spreading from Texas to Pennsylvania, Colorado, North Dakota and other parts of the United States (see Burwen and Flegal 2013).

Although fracking remains largely a U.S. phenomenon, it has become a major, and contentious, issue in other parts of the world. Fracking continues to be a key driver in the geopolitics of energy (EIA 2013; Blackwill and O’Sullivan 2014). As just one important example of this, the United States has rapidly gone from plans to import liquefied natural gas (due to domestic scarcity) to exporting it (due to domestic oversupply). Once these exports come online, they will drastically alter energy markets, with major political ramifications.

The Science Politics of Fracking

Fracking is one of the most divisive technologies of our time. A good deal of the controversy hinges on the old chestnut in the philosophy of technology about whether certain social conditions

must inevitably follow from the use of an artifact, or whether society itself can shape the artifact in different ways to yield different outcomes. To wit, some claim that “fracking is inherently dangerous” such that no rules could make it acceptable (Steingraber 2012) while others argue that fracking can be safe as long as the “highest practicable standards” are applied (IEA 2012).

The controversies surrounding fracking are reminiscent of debates about human cloning or, more analogously, nuclear power. Indeed, the conflicts over fracking possess religious undertones not unlike the anger and hatred of the *odium theologicum* wherein incommensurable theological views are pitted against one another. For some, fracking is the savior. For others, it is the devil. One side says: fracking creates economic growth and jobs, lowers carbon emissions, bolsters energy security, and does not harm air and water. The other side says: fracking exacerbates climate change, ruins local communities, serves primarily the interests of multinational corporations, and pollutes air and water.

The expectation on both sides is that science will resolve the controversies by revealing who is right and who is wrong. Indeed, both sides appeal to science in making their case for and against fracking. Both wave the same banner: *Wissenschaft mit uns!* (Science on our side!) But can science support both the pro and con position?

The answer from nearly everyone within the opposing camps is “no.” What’s really going on, they say, is that their side has the real science while the other side has junk science. Thus, the basic dynamic in the feud is one of debunking or unmasking: showing how the other side is falsifying, fabricating, or otherwise distorting the science to support foregone, and false, conclusions.

Josh Fox, maker of the anti-fracking documentary *Gasland*, thematized this most explicitly in his short film *The Sky Is Pink*. His central claim is that the oil and gas industry manipulates the media to create doubt where there is in fact scientific certainty. By prolonging uncertainty, the industry is able to stave off regulations and other threats to its business model. Naomi Oreskes and Eric Conway (2011) dub this the “tobacco strategy,” because cigarette companies were the first to utilize a systematic campaign of prevarication and misinformation to stall regulations on a harmful product. Much the same, they argue, is occurring with climate change. With fracking, this often takes shape around claims of ‘frackademia,’ where the oil and gas industry surreptitiously corrodes the autonomy of university research. The aim is to cloak its interests in the guise of independent science, thereby giving its claims the credibility and imprimatur of the academy (for one example, see Coleman 2015).

But such accusations cut both ways: Oil and gas industry proponents similarly accuse anti-fracking groups of using junk science with undisclosed ties to biased sources. For example, the fracking public relations group Energy In Depth has written extensively about the Park Foundation as a secretive source of funding for “an army of anti-fracking activists.” Writing for Energy In Depth, Tom Shepstone (2012) argues that a *New York Times* series of articles that was very critical of fracking was sourced in large measure from people and organizations supported by the Park Foundation but that that connection was not disclosed in the series. More generally, industry groups and their proponents tend to claim that opposition to fracking is primarily driven by irrational fears and ignorance, such that once people see “the facts” they will come around to a more positive assessment of fracking (for example, Macdonald 2015).

Yet it is also possible that the answer to the question of whether science can support both sides is “yes.” Take the issue of water contamination, arguably the most important and contentious debate related to fracking, as an example. The concern here is that the chemicals used (and radioactive and other harmful elements that exist naturally deep underground) can be introduced into water supplies through hydraulic fracturing, well casing failures, underground wastewater injection, or above-ground spills via leaking tanks, trucks, or pits (see Lustgarten 2008; Colborn et al. 2011; EPA 2012).

In 2015, the U.S. Environmental Protection Agency released a long-anticipated (and long-delayed) study of this issue (EPA 2015). Tellingly, both sides immediately marshaled the study as further ammunition in their war. Writing for *Energy In Depth*, Katie Brown (2015) argued that the EPA study exonerated the industry and debunked anti-fracking activists. She emphasized the language from the EPA that stated that it found “no widespread, systemic impacts” to groundwater. By contrast, Fox interpreted the study in the opposite light as confirmation that fracking does indeed pollute groundwater, which debunks the industry’s claims that there has never been a single proven case of groundwater contamination (Fox and Ziesche 2015).

What’s going on here?

Part of the issue is definitional in nature. A great deal of the groundwater contamination debate appears like ships passing in the night, because the two sides are defining ‘fracking’ in different ways. The industry tends to use a narrow definition limited to the actual injection of water and chemicals thousands of feet below ground. Fracking opponents tend to use a broader definition that encompasses the whole drilling and production process.

More importantly, as Daniel Sarewitz (2004) argues, there is “an excess of objectivity” such that both sides will always be able to find a body of facts to support their position. The natural-technical-social systems are sufficiently complex and science itself is sufficiently diverse that no “one right way” of seeing the situation exists.

A good illustration of this is the debate surrounding the air quality and climate impacts of fracking. Sources of emissions from fracking, understood broadly, include leaks (from pipelines, valves, or condensate tanks), routine venting, on-site engines to operate equipment, and the truck engines that supply materials. The types of emissions include methane (CH_4), ozone precursors such as Volatile Organic Compounds (VOCs) and nitrogen oxides (NO_x), and Hazardous Air Pollutants (HAPs) such as benzene and formaldehyde (Alvarez and Paranhos 2012). There is disagreement about the net impact of fracking on air quality, especially as natural gas produced via fracking displaces coal as a resource for electricity generation (compare, for example, the conclusions of Armendariz 2009 and McKenzie et al. 2012 with Ireland 2009 and TCEQ 2011). Here too, this disagreement is conducted through a battle of competing scientific studies.

A similar controversy exists over the climate impacts of fracking (see EPA 2013), with proponents crediting fracking for reducing U.S. CO_2 emissions (see Lomborg 2012) and opponents arguing that, largely due to fugitive emissions of methane, the greenhouse-gas footprint of shale gas equals or exceeds that of coal (Howarth et al. 2011). Each side has a body of knowledge to bolster its claims, such that the debate hinges on which kinds of methods, models, and measurements are most suitable. On top of this debate, proponents also argue that the success of fracking has bolstered investments in renewable energy technologies that can work in tandem with natural gas (billed as a ‘bridge fuel’), while opponents argue that fracking has depressed the cost of oil and gas, removing the key incentive to invest in renewable technologies.

One could argue that such debates will not be resolved by sifting the junk science from the real science, because legitimate kinds of scientific evidence support conflicting claims. The differences are about which facts are emphasized and how they are framed within a larger narrative concerning values. How much risk is acceptable? How do we weigh competing goods and rights claims? What kind of evidence is most relevant to the policy at hand? What is the proper place of humans in nature? What kind of world do we want to live in and pass down to our children? These questions are not reducible to science. There is no scientific way to determine an optimal distance between fracking sites and playgrounds. That decision depends on how we want to balance the goals of safety, community character, and access to mineral rights. Indeed, there is no scientific answer to the question of how much scientific research to perform and which scientific findings to emphasize.

The controversy about the economic impacts of fracking further highlights the way research is mobilized to support conflicting views. A 2011 report prepared for America's Natural Gas Alliance claimed that shale gas in the United States supported 600,000 jobs and added \$76 billion to the GDP (IHS 2011). Other regional reports also claim major economic benefits from fracking (e.g., Perryman Group 2011). Such economic impact assessments, however, require numerous and often ethically charged assumptions, especially about long-term impacts and which costs (e.g., harms to the environment or public health) are included (Christopherson and Rightor 2011). Indeed, disputes about the economics of fracking are often about whether and how to account for negative externalities (uncompensated harms or costs not reflected in the price of oil or gas). They can also be seen as debates about environmental justice (see Fry et al. 2015).

I can draw from my own experience with fracking to illustrate these points. I worked for years with the city of Denton, Texas, to craft local regulations on fracking. Due to a variety of factors I detail in my book on this subject (Briggle 2015), our local regulations proved to be ineffective. At that point, I helped to lead an effort to ban the use of hydraulic fracturing in the city limits. During the campaign before the vote on the proposed ban, the industry argued that the ban would harm the local economy. It commissioned a study, which concluded that the ban would cost the city \$251 million in gross product (Perryman Group 2014). I was able to use the very same numbers in that report and put them in a different light (Briggle 2014). For example, its own numbers showed that the ban would impact 0.5% of the city's tax revenues, 0.25% of the workforce, and 0.17% of the school district budget. I then showed how the ban would actually generate more tax revenues for the city, because homes create far more economic growth than frack sites.

The point isn't to say that the report is 'junk science,' but rather that numbers can be interpreted and presented in different ways. The same is true with concerns about how much water fracking consumes. Seen from one perspective, it is millions of gallons of water permanently disposed of, often in water-stressed areas (Freyman and Salmon 2013). Seen from another perspective, it is only 0.04% of the total freshwater use in the United States (Kondash and Vengosh 2015).

Of course, protecting the integrity of science and disclosing potential sources of bias are extremely important, because science is needed to guide policy. But it is never going to be sufficient to *determine* policy. The link between fracking and earthquakes is a good case in point. This is one instance where the science has generally been settled: on several occasions, the U.S. Geological Survey has definitively linked disposal wells (that are used to store waste water from oil and gas production) with induced earthquakes. Though the oil and gas industry, and state regulatory agencies heavily influenced by the industry, first resisted this conclusion, there is now fairly widespread agreement (Wines 2015).

But this has done little to resolve any policy questions, let alone convert basic attitudes about fracking. After all, of the 35,000 active injection wells in the United States, only a few dozen have been shown to cause felt earthquakes. And much of the wastewater dumped down disposal wells is not (partly depending on the definition used) related to fracking (see Rubinstein and Mahani 2015). There are, then, a wide range of viable alternative courses of action that are consistent with the science. So, even when there is conclusive scientific evidence of cause and effect relationships, policy disputes live on.

Real-World Experimentation

One of the most salient criticisms of modernity is that our capacity for action has outstripped our capacity for thought. In *The Communist Manifesto*, Karl Marx wrote that modern bourgeois

society is like a magician who has conjured up such powerful forces that he is no longer able to control them. A hundred years later, Günther Anders argued that *herstellen* (what we can produce) has outgrown *vorstellen* (what we can conceive). Hannah Arendt wrote about the “world alienation” that ensues when we can no longer make sense of human experience (Arendt 1958).

Of course, in modern society there is a general presumption in favor of action, which usually goes by the term *innovation*. The new is *prima facie* welcomed and considered better than the old. Nonetheless, especially as human powers magnify, there is an uneasy sense that perhaps we have not thought things through. It could be that further technological activity creates unanticipated harms to go along with the expected benefits. What should we do about the fact that freedom to pursue intended desired outcomes (like cheap energy) also creates unintentional harms (environmental or otherwise)?

Precaution is often offered as one *principled* answer to this question: We need to first stop and think prior to acting. The technology is treated as guilty until proven innocent. The guiding principle is to take more factors into account prior to enrolling technologies into society—consider not just the narrow intention but the wider ramifications.

By implementing a six-year moratorium on hydraulic fracturing while conducting studies about potential health, environmental, and economic impacts, the state of New York adopted a precautionary stance. In late 2014, New York prohibited hydraulic fracturing on the basis of these studies, with the health commissioner claiming that they uncovered “significant public health risks.” He added that “The potential risks are too great. In fact, they are not even fully known” (in Kaplan 2014).

Yet, as the preceding discussion suggests, there is no way to “fully know” risks. A technology cannot be proven innocent, especially before it has been widely deployed. To demand certainty about broader impacts is to postpone action indefinitely. This is something that Jacques Ellul pointed out in *The Technological Order*:

But there are always secondary effects which had not been anticipated, which in the primary stage of the technical progress in question could not in principle have been anticipated. This unpredictability arises from the fact that predictability implies complete possibility of experimenting in every sphere, an inconceivable state of affairs.

(in Mitcham and Mackey 1983: 104)

So, if we are going to get the intended benefits of a technology such as fracking, we must also risk getting unintended harms. At some point, the experiment must move out of the lab and into the world. It could be claimed that citizenship in the modern world is the same as being a research subject in a panoply of engineering experiments (see Martin and Schinzinger 1989).

This condition is itself generally un-thought—it is simply lived out as the given state of affairs. Indeed, one reason why precaution is discussed as a *principle* is because it stands as the product of conscious reflection in distinction to the default real-world experimentation of modern existence. Yet there are thoughtful articulations of this condition, including Karl Popper’s notion of “piecemeal social engineering.” Popper (1957) thought an open society made small-scale, incremental changes in light of ever-evolving knowledge and experience rather than large-scale changes based on some grand narrative or presumption that we have it all figured out in advance of taking action in the world. To innovate, learn, revise, and adapt is the kind of evolutionary process that characterizes both democratic and scientific progress.

A more recent formulation was offered by the transhumanist Max More in terms of the “proactionary principle” (More 2005). More argues that the precautionary principle stifles human freedom and progress: had it been applied in the past we would not now have electricity, pathogen-free water, X-rays, vaccinations, and much more. “Being proactive,” he claims,

“involves not only anticipating *before* acting, but learning *by* acting.” He concludes triumphantly: “Let a thousand flowers bloom! By all means, inspect the flowers for signs of infestation and weed as necessary. But don’t cut off the hands of those who spread the seeds of the future” (More 2005: np).

More’s principle contains three criteria for evaluating real-world experiments. For a real-world experiment to be just or ethical:

1. Those most vulnerable to the unintended harms must give their consent to the risk or, at the very least, they must be compensated for any harms incurred.
2. There must be a robust system for monitoring and learning from the real-world experiment.
3. The experiment must be modifiable when problems are identified through the learning process, that is, the original innovation must be readily renovated.

Elsewhere, I have used these criteria to assess fracking, especially as it has taken shape in the context of Denton, Texas (Briggles 2015). I argued that there are several areas in need of reform in order to bring common fracking practices in line with these criteria. For example, the use of nondisclosed, proprietary chemicals in fracking undermines the conditions necessary both for consent (because people are not adequately informed) and compensation (because secrecy hampers the ability to trace causal connections). As another example, the use of nondisclosure clauses in lease agreements hampers the ability of society to learn from past mistakes.

More is ambiguous when it comes to the relationship between precaution and proaction. Mostly, he sees them as opposites. But he also suggests that they can be complementary—that we can *both* think before acting and think while acting. This did not occur in Texas, where the contemporary constellation of fracking was born—there was no systemic thought about wider unintended impacts prior to massive commercialization of the technology. Yet arguably this kind of thinking has occurred in the United Kingdom, which has also cautiously begun to act by issuing permits for onshore fracking operations (although this activity remains highly contentious).

Not surprisingly, fracking opponents tend to favor precautionary approaches that would prohibit fracking until it can be proven safe (understanding that such proof is impossible). Although it may be easy to dismiss this political tactic as self-serving, there is an important—indeed potentially damning—criticism of those who favor a more proactionary approach of learning and adjusting while acting: proving the guilt of fracking is also impossible. Once a technology is enrolled in society (especially within a neoliberal capitalist order), powerful actors have vested interests in *not* finding anything amiss with the real-world experiment. The bar for establishing ‘proof’ can be moved ever higher and the evidence presented can be forever scrutinized.

Thus, some of the most important questions about fracking are: who bears the burden of proof, and what are the standards for establishing sufficient evidence of guilt or innocence? Too often in the United States it is average citizens with little access to institutional resources and credible epistemic authorities who must establish evidence of harm to satisfy very exacting legal and scientific standards.

Perhaps the biggest question is not *what* the rules governing fracking should be, but *who* is empowered to write the rules. This invites very complex legal and ethical considerations about jurisdiction, because fracking is simultaneously a matter of energy policy (thus the business of state and federal governments) and land use policy (thus the business of local governments). Atop this dynamic is layered the many complexities of private property ownership, which entails the participation of individuals and corporations with these various levels of government.

In most countries, mineral wealth is owned by the central government, which at least gives citizens some indirect, representative voice in determining whether and under what conditions

leases are offered to either state-owned or privately held companies. In the United States, mineral wealth is privately owned, and it is often the case that the surface property is severed from the subsurface property. That creates conditions where those living on the surface near fracking sites are exposed to the harms but do not receive any direct financial compensation and are excluded from the decision-making process about leases. This is why local governments (and the issue of state preemption of local government) have become so central in the controversies about fracking. Citizens who are excluded from private decisions on the market can find a public voice at their city hall when it comes to participating in decisions that impact their lives.

Conclusion

To conclude, it is worth putting—even if only in a wryly ironic way—the experiment that is fracking into perspective. Certainly, it is of enormous significance in shaping fortunes, health, power, and the planet. But there is perhaps nothing all that new about it. Martin Heidegger might note that it is a continuation of a basically extractivist orientation toward nature. Karl Marx might note that it still unfolds within the structures of a stratified, property-based, global capitalist order. And Aristotle might note that it is all still in the service of a hedonistic or utilitarian conception of the good life.

Indeed, for as vehemently as they clash, both sides in the fracking debate subscribe to a “type 1 energy ethics,” which “holds energy production and use as a fundamental good and is grounded in the assumption that increases in energy production and consumption result in increases in human wellbeing” (Mitcham and Rolston 2013: 313). From this perspective, the differences are superficially about the *means* for achieving the good life (whether via oil and gas or solar and wind power). Very few in the fracking debate take seriously an alternative vision of human flourishing premised on lower energy consumption.

A “type 2 energy ethics” questions the linear relationship between energy and welfare. Ivan Illich (1973) noted how *all* advanced energy systems tend to alienate citizens, turning them into consumers of goods and services provided by systems managed by experts. The resulting technocracy breeds dependence and passivity, which undermine human equity, autonomy, and happiness.

The debate over fracking would be broadened and enriched were it to take seriously a type 2 energy ethics. However, it is part of the character of our age that our experimental spirit is applied primarily to the world rather than to our souls. It would stand to reason that an age that would go to such extremes to supply itself with energy must be putting that energy into the service of something noble and even transcendent. Yet there is little evidence that our hyper-active culture is superior to others, unless we contain our measurement within the confines of instrumentality and speak only of our superior *means*. And sadly, there is a great deal of evidence to suggest that we have lost touch with the transcendent and have chosen longevity over nobility.

We know how to frack, but we know not why.

References

- Alvarez, R., and Paranhos, E. (2012). Air Pollution Issues Associated with Natural Gas and Oil Operations. *Air and Waste Management Association*, June, 22–25.
- Anders, Gunther. (1972). *Endzeit und Zeitende: Gedanken über die atomare Situation*. Beck: München.
- Arendt, H. (1958). *The Human Condition*. Chicago: University of Chicago Press.
- Armendariz, A. (2009). *Emissions From Natural Gas Production in the Barnett Shale Area and Opportunities for Cost-Effective Improvements*. Available at: www.edf.org/sites/default/files/9235_Barnett_Shale_Report.pdf.

- Blackwill, R. D., and O'Sullivan, M. L. (2014). America's Energy Edge: The Geopolitical Consequences of the Shale Revolution. *Foreign Affairs*, March/April. Available at: www.foreignaffairs.com/articles/140750/robert-d-blackwill-and-meghan-l-osullivan/americas-energy-edge.
- Briggle, A. (2014). Fracked on Their Own Petard. *Truthout*, August 11. Available at: www.truth-out.org/opinion/item/25429-fracked-on-their-own-petard-the-self-implosion-of-an-industry-on-the-ropes#.
- Briggle, A. (2015). *A Field Philosopher's Guide to Fracking*. New York: Liveright.
- Brown, K. (2015). Ten Important Things to Know From EPA's 1,000-page Groundwater Study, June 11. Available at: <http://energyindepth.org/national/ten-important-things-to-know-from-epas-1000-page-groundwater-study/>
- Burwen, J., and Flegal, J. (2013). Unconventional Gas Exploration and Production. *American Energy Innovation Council*, 1–16. Available at: <http://americanenergyinnovation.org/wp-content/uploads/2013/03/Case-Unconventional-Gas.pdf>.
- Carlisle, N., and Carlisle, J. (1967). Project Gasbuggy. *Popular Mechanics*, 128, September, 104, 105, 222.
- Christopherson, S., and Rightor, N. (2011). A Comprehensive Economic Impact Analysis of Natural Gas Extraction in the Marcellus Shale. *Cornell University Working Paper Series*, May. Available at: www.greenchoices.cornell.edu/downloads/development/marcellus/Marcellus_SC_NR.pdf.
- Colborn, T., Kwiatkowski, C., Schultz, K., and Bachran, M. (2011). Natural Gas Operations From a Public Health Perspective. *Human and Ecological Risk Assessment*, 17(5), 1039–1056.
- Coleman, J. (2015). University of Colorado Part of Fracking PR Scheme. *The Huffington Post*, September 17. Available at: www.huffingtonpost.com/jesse-coleman/university-of-colorado-pa_b_8155404.html.
- DOE (U.S. Department of Energy) (2007). *DOE's Unconventional Gas Research Programs 1976–1995: An Archive of Important Results*, January 2007, Available at: www.netl.doe.gov/kmd/cds/disk7/disk2/Final%20Report.pdf.
- DOE (U.S. Department of Energy) (2009). *Modern Shale Gas Development in the United States: A Primer*, April. Available at: www.netl.doe.gov/technologies/oil-gas/publications/EPreports/Shale_Gas_Primer_2009.pdf.
- EIA (U.S. Energy Information Administration) (1993). *Drilling Sideways—A Review of Horizontal Well Technology and Its Domestic Application*, April. Available at: www.eia.gov/pub/oil_gas/natural_gas/analysis_publications/drilling_sideways_well_technology/pdf/tr0565.pdf.
- EIA (U.S. Energy Information Administration) (2013). *Annual Energy Outlook 2013 With Projections to 2040*, April. Available at: [www.eia.gov/forecasts/aeo/pdf/0383\(2013\)pdf](http://www.eia.gov/forecasts/aeo/pdf/0383(2013)pdf).
- EPA (U.S. Environmental Protection Agency) (2010). *Hydraulic Fracturing Research Study*, June. Available at: www.epa.gov/safewater/uic/pdfs/hfresearchstudyfs.pdf.
- EPA (U.S. Environmental Protection Agency) (2012). *The Potential Impacts of Hydraulic Fracturing on Drinking Water Resources: Progress Report*, December. Available at: <http://www2.epa.gov/hfstudy/potential-impacts-hydraulic-fracturing-drinking-water-resources-progress-report-december>.
- EPA (U.S. Environmental Protection Agency) (2013). *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2011*, April. Available at: www.epa.gov/climatechange/Downloads/ghgemissions/U.S.-GHG-Inventory-2013-Main-Text.pdf.
- EPA (U.S. Environmental Protection Agency) (2015). *Assessment of the Potential Impacts of Hydraulic Fracturing for Oil and Gas on Drinking Water Resources (External Review Draft)*, June. Available at: <http://cfpub.epa.gov/ncea/hfstudy/recordisplay.cfm?deid=244651>.
- Fox, J., and Ziesche, L. (2015). Fracking Does Cause 'Widespread, Systemic' Contamination of American's Drinking Water. *EcoWatch*, June 5.
- Freyman, M., and Salmon, R. (2013). Hydraulic Fracturing and Water Stress: Growing Competitive Pressures for Water. *Ceres*, May. Available at: www.ceres.org/resources/reports/hydraulic-fracturing-water-stress-growing-competitive-pressures-for-water.
- Fry, M., Briggle, A., and Kincaid, J. (2015). Fracking and Environmental (In)justice in a Texas City. *Ecological Economics*, 117, 97–107.
- Howarth, R., Santoro, R., and Ingraffea, A. (2011). Methane and the Greenhouse-Gas Footprint of Natural Gas from Shale Formations. *Climatic Change*, 106(4), June, 679–690.

- IEA (International Energy Agency) (2012). *Golden Rules for a Golden Age of Gas*, May. Available at: www.worldenergyoutlook.org/goldenrules/.
- IHS (2011). *The Economic and Employment Contributions of Shale Gas in the United States*, December. Available at: www.ihs.com/info/ecc/a/shale-gas-jobs-report.aspx.
- Illich, I. (1973). *Energy and Equity*. New York: Harper.
- Ireland (ed.) (2009). *Barnett Shale Natural Gas Production and Air Quality*. Available at: www.barnettshalenews.com/documents/ntxairstudy/Review%20of%20Armendariz%20Study%20of%20Barnett%20Shale%20Pollution%20by%20Dr%20Ed%20Ireland%203-11-2009.pdf.
- Kaplan, T. (2014). Citing Health Risks, Cuomo Bans Fracking in New York State. *New York Times*, December 17. Available at: www.nytimes.com/2014/12/18/nyregion/cuomo-to-ban-fracking-in-new-york-state-citing-health-risks.html?_r=0.
- Kondash, A., and Vengosh, A. (2015). Water Footprint of Hydraulic Fracturing. *Environmental Science and Technology Letters*, 2, 276–280.
- Lomborg, B. (2012). A Fracking Good Story. *Project Syndicate*, September 13. Available at: www.project-syndicate.org/commentary/a-fracking-good-story-by-bj-rn-lomborg.
- Lustgarten, A. (2008). Buried Secrets: Is Natural Gas Drilling Endangering U.S. Water Supplies? *Pro Publica*, November 13. Available at: www.propublica.org/article/buried-secrets-is-natural-gas-drilling-endangering-us-water-supplies-1113.
- MacDonald, A. (2015). It's True Women Don't Like Fracking. I Want to Change That. *The Guardian*, October 23.
- Marx, K., Engels, F., Moore, S., & McLellan, D. (1992). *The Communist manifesto*. Oxford: Oxford University Press.
- Martin, M., and Schinzing, R. (1989). *Ethics in Engineering*, 2nd edition. New York: McGraw-Hill.
- McKenzie, L., Witter, R., Newman, L., and Adgate, J. (2012). Human Health Risk Assessment of Air Emissions From Development of Unconventional Natural Gas Resources. *Science of the Total Environment*, 424, 79–87.
- Mitcham, C., and Mackey, R. (eds.) (1983). *Philosophy and Technology*. New York: Free Press.
- Mitcham, C., and Rolston, J. S. (2013). Energy Constraints. *Science and Engineering Ethics*, 19, 313–319.
- Montgomery, C., and Smith, M. (2010). Hydraulic Fracturing: History of an Enduring Technology. *Journal of Petroleum Technology*, December, 26–41.
- More, M. (2005). The Proactionary Principle. Available at: www.maxmore.com/proactionary.htm.
- Oreskes, N., and Conway, E. (2011). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming*. New York: Bloomsbury.
- Perryman Group (2011). *A Decade of Drilling: The Impact of the Barnett Shale on Business Activity in the Surrounding Region and Texas: An Assessment of the First Decade of Extensive Development*, August. Available at: www.fortworthchamber.com/am-site/media/barnett-shale-study.pdf.
- Perryman Group (2014). *The Adverse Impact of Banning Hydraulic Fracturing in the City of Denton on Business Activity and Tax Receipts in the City and State*, June. Available at: <http://energyindepth.org/wp-content/uploads/2014/07/Perryman-Denton-Fracking-Ban-Impact.pdf>.
- Popper, K. (1957). *The Poverty of Historicism*. New York: Routledge and Kegan Paul.
- Rubinstein, J., and Mahani, A. (2015). Myths and Facts on Wastewater Injection, Hydraulic Fracturing, Enhanced Oil Recovery, and Induced Seismicity. 86(4), July/August, 1–8. Available at: https://profile.usgs.gov/myscience/upload_folder/ci2015Jun1012005755600Induced_EQs_Review.pdf.
- Sarewitz, D. (2004). How Science Makes Environmental Controversies Worse. *Environmental Science & Policy*, 7, 385–403.
- Shepstone, T. (2012). Connecting the Dots on NYT and Park Foundation. *Energy In Depth*, February 14. Available at: <http://energyindepth.org/marcellus/connecting-dots-park-foundation/>.
- Steingraber, S. (2012). Safe Hydrofracking is the New Jumbo Shrimp. *Huffington Post*, June 4. Available at: www.huffingtonpost.com/sandra-steingraber/safe-hydrofracking_b_1520574.html.
- Steward, D. (2007). *The Barnett Shale Play: Phoenix of the Fort Worth Basin, a History*. Fort Worth: Fort Worth Geological Society Press.
- TCEQ (Texas Commission on Environmental Quality) (2011). *A Commitment to Air Quality in the Barnett Shale*. Austin, TX: Texas Commission on Environmental Quality. Available at: <http://tinyurl.com/3mc5l6c>.

-
- Wang, Z., and Krupnick, A. (2013). A Retrospective Review of Shale Gas Development in the United States. *Resources for the Future*. Available at: www.rff.org/RFF/documents/RFF-DP-13-12.pdf.
- Wines, M. (2015). Oklahoma Recognizes Role of Drilling in Earthquakes. *New York Times*, April 21. Available at: www.nytimes.com/2015/04/22/us/oklahoma-acknowledges-wastewater-from-oil-and-gas-wells-as-major-cause-of-quakes.html.
- Yost, A. B., and Overbey Jr., W. K. (1989). Production and Stimulation Analysis of Multiple Hydraulic Fracturing of a 2,000-ft Horizontal Well. *Society of Petroleum Engineers Gas Technology Symposium*, June 7–9.

Climate Change and Philosophy of Technology

Robert-Jan Geerts

Introduction

Without the technologies that we use every day, we would not be able to change the climate in the way that we currently are. But despite the radically technological nature of climate change, discussions have thus far benefitted only sporadically from insights from philosophy of technology. Why is this, and how can we change this? This chapter will tackle these questions in three steps. First, it gives a brief overview of the general avenues along which climate change is typically discussed, and makes some suggestions for how philosophy of technology may inform those discourses. Second, it sketches some alternative routes to understand climate change starting with perspectives emerged in philosophy of technology. And finally, it turns the problem around and points out how climate change lays bare a problem with recent trends in philosophy of technology.

Mainstream Climate Change Discourse

Anthropogenic climate change is understood as one of the most important global challenges faced by current society. The emission of greenhouse gases (most notably carbon dioxide, or CO₂) and land use changes such as deforestation are, according to many scientists, causing changes in the global climate. The extent and exact distribution of these changes are unclear, but they are likely to involve an increase in average temperature, changing precipitation patterns, and an increase in frequency and intensity of extreme weather events such as hurricanes, floods, and droughts. These changes are already measurable, and are expected to become much more severe over the next decades and into the distant future. Climate change threatens people's safety (due to extreme weather events), habitat (most notably in coastal low-lying areas and islands), agricultural practices, and clean water supply. It also puts individual animals, biological species, natural areas, and fragile ecosystems at risk.

Discussions on climate change and what to do about it are diverse. General introductions as well as proposed routes to solutions are widely available, but plenty of controversy remains. In this section, a number of perspectives and central discussion topics are introduced.

Intergovernmental Panel on Climate Change (IPCC)

The most prestigious authority on the matter of climate change is the IPCC, an assembly of thousands of scientists, technologists, and policy makers under the flag of the United Nations. It assesses and publishes policy-relevant reports on developments in climate science (e.g., IPCC 2014). The IPCC is divided into three working groups, focusing on different aspects of the

climate change problem. By splitting up the discussion, governments can officially endorse the findings of one working group, without subscribing to the conclusions of another.

Working Group I focuses on the scientific understanding of climate change. It collects and assesses research outcomes concerning change in temperature, precipitation patterns, extreme weather events (e.g., droughts, floods, hurricanes), snow and ice cover, and sea level, at both global and regional levels. Of these elements of the global climate, it assesses what changes have already occurred when compared to the preindustrial situation, whether and how the changes are anthropogenic (caused by humans), and how these elements are expected to develop further, depending on a number of future scenarios for greenhouse gas emissions.

Working Group II focuses on adaptation to climate change. This working group looks at climate risks (split into regions), proposes ways to adapt to the climate change and to reduce the accompanying risks, and assesses to what extent this risk reduction is at all possible. The assessment is split between current climate change and a number of future scenarios. Which of these scenarios will turn out to be realistic depends on uncertainties in the climate science and to what extent future greenhouse gas emissions will be mitigated. The working group suggests that many risks can be reduced, but not completely eradicated. Some, like the disintegration of coral reefs, are likely to happen anyway.

Working Group III focuses on climate change mitigation. It first sketches which activities are actually causing climate change, split into energy production; industry; transport; buildings; and agriculture, forestry, and other land use (AFOLU). For each of these sectors, mitigation solutions are proposed. It is emphasized that mitigation needs to be a cooperative effort, that mitigation choices typically involve ethical considerations, and that many mitigation efforts also benefit other societal goals. Mitigation scenarios are built around pathways towards certain CO₂ concentrations (450–1,000 ppm) in 2100, which are connected to an expected average warming since the preindustrial situation, between 1.5 and 4.8 degrees Celsius. The sooner changes are made, the cheaper they will be, and the more feasible more ambitious targets remain. This is especially true for infrastructure investments, due to their long lifetime and lock-in effects (for example, cities built for the automobile mean people will have to keep driving their cars). The working group suggests efficiency improvements, behavior change, and subsidizing of novel technologies (most notably renewable energy projects) as key components of mitigation strategies.

The IPCC mitigation strategies are situated within a paradigm of continued development. Population growth, urbanization, and economic growth are more or less assumed to be givens; mitigation efforts are to reduce climate change while continuing on the path that society has set for itself. ‘Co-benefits’ are emphasized: we can improve the climate *and* these other things governments are trying to achieve. This position is not surprising for a body that needs to inform and coordinate global and national climate policy, but it is not unproblematic. In the next section, some critiques of this position will be discussed.

Beyond the work of the IPCC, the mainstream of climate debates may be organized between political and technological commentaries, even though both of these may have elements of the other in them. Perhaps it is useful to understand them as a spectrum, although it seems most of the time they are written clearly from one perspective.

Political Discourse on Climate Change

The political discourses on climate change revolve around knotty philosophical and ethical questions that emerge when trying to react to the threats of climate change. A good place to start is Stephen Gardiner’s *A Perfect Moral Storm* (2006), in which climate change is understood as

a particularly tricky version of a tragedy of the commons. This ethical puzzle was first described by Garrett Hardin (1968) and describes a situation in which individual rational actors make decisions that are independently beneficial to the actors themselves, but in the aggregate detrimental to the group, and ultimately to the actors themselves. The paradigm example is of a pasture shared by several shepherds. By adding more sheep to a private herd a shepherd gains income, while putting only marginally more pressure on the pasture. But when all shepherds act this way, overgrazing will occur and the pasture will eventually collapse. To solve this problem, a system of “mutual coercion, mutually agreed upon” (Hardin 1968: 1245) is suggested: for their own good, the shepherds must all accept the limits of the pasture.

According to Gardiner, we must understand the atmosphere’s ability to absorb greenhouse gases safely in exactly this way, except that the traditional solution of “mutual coercion, mutually agreed upon” is difficult to achieve in this case. The problem is threefold. First, this commons is global, which means that actors from different cultures in a vastly unequal environment need to come to an agreement. Second, the commons is intergenerational. If actors from all over the world can at least in theory get together and discuss, actors from different generations cannot. What is more, the current generation is the only one able to act now, but it is also the generation that benefits the least from responsible behavior, because the worst effects of climate change occur in the future. The third problem is that, according to Gardiner, we lack theoretical tools and frameworks to deal effectively with issues at the core of climate change, such as scientific uncertainty, long-term future planning, and environmental ethics, for example. Together, these three issues result in a situation in which we are very susceptible to moral corruption. We may add to this susceptibility the central position of fossil energy and its benefits to modern society: it is most certainly something substantial to let go.

A specific issue within the political discourse on climate change is the notion of *climate justice*. If one accepts that climate policy involves the burden of a certain restraint, then what follows is the question of how to distribute this burden: in addition to being effective, climate policies need to be just. There are many ways to go about this. Jeremy Baskin (2009), for example, distinguishes four routes: equal cuts for all, converge and contract, greenhouse gas development rights, and geoengineering. Equal cuts for all is understood as unjust because it leads to a relative status quo: if everyone cuts emissions by 70%, the biggest polluters now will remain large polluters, while the poor have to cut their already tiny emissions to a level at which it could very well be impossible to survive. The converge-and-contract approach resolves these problems by aiming for an equal emission level for everyone at some point in the future. This means the largest polluters have to cut their emissions most, and only the poorest nations are able to increase their emissions slightly. Still, this is problematic because the West has access to more advanced technologies that allow for higher standards of living with the same emission levels: it would be unfair, Baskin argues, to not let developing countries develop towards a similar technological level; however, an infrastructure of highways and airports that imply high-emission lifestyles in the West make the necessary large emission cuts very difficult to achieve. A step even further in this direction is the greenhouse gas development rights framework, which holds that everyone should be able to develop to a similar level of affluence before paying for mitigation efforts. The result of such a framework would be that the developed West would pay for further development of poor nations, leading to an even heavier burden on the West. The final approach, geoengineering, amounts to a technological fix that is as yet shrouded in uncertainty, but that would most likely result in a mix of ‘winners’ and ‘losers.’ Geoengineering is discussed in more detail in the final section of this chapter.

Baskin points out that, although difficult to obtain, justice is an essential part of a successful climate change mitigation approach. But using many of the same arguments, other scholars are led in the opposite direction. In *Climate Change Justice*, Posner and Weisbach (2010) argue

that exactly because we need global cooperation, climate policy should not be used to solve all kinds of other issues. Instead, we should work on policies that are based on the principle of *international paretianism*: each party must believe that they are better off with a climate treaty than without it. Climate policy would thus be acceptable only when for each party it involves lower costs than benefits. It is obvious that there is little room for the considerations put forth by Baskin when each nation puts its own demands first.

These considerations are central to debates during UN Climate Change Conferences, annual meetings of parties to the Framework Convention on Climate Change (UNFCCC), first negotiated at the 1992 'Earth Summit' in Rio de Janeiro. The 1997 *Kyoto Protocol* focused only on emission reductions by developed nations, effectively subscribing to a converge-and-contract approach. The 2015 *Paris Agreement* included the pledge of developed nations to fund climate mitigation and adaptation projects in developing nations with a value of at least US \$100 billion per year, suggesting subscription to something along the lines of a greenhouse gas development rights framework. However, as the agreement does not include strict enforcement, actual commitment to these pledges remains in the hands of national governments, which may have difficulty deploying policies that are perceived as going against the interests of the nation.

Technological Discourse on Climate Change

Perhaps the most widely read technical approach to climate change mitigation is *Sustainable Energy—Without the Hot Air* (MacKay 2009). Although sustainable energy does not exactly coincide with climate change mitigation (there are other reasons beyond climate change to pursue a sustainable energy system and more than just energy involved in climate change), the two overlap enough to understand discussions of one as central to the other. In his book, MacKay argues that the current discussion regarding energy consumption and emissions is unnecessarily troubled because of a bad choice of units. Claiming a new wind farm provides electricity for 5,000 households does not provide any insight in its significance for the world. Therefore, MacKay suggests we should look at the emissions per person. Only then it becomes clear how our emissions are distributed, so we can see where the big savings are (stop flying), and what does not add much (unplugging our telephone chargers).

MacKay then proceeds with showing that with a smart application of novel and proven technologies, a little mentality change, and some dedication, it is indeed possible to make the transition towards a climate-neutral society. He also shows how much of Britain's surface area needs to be used for these novel technologies, and what technological problems are still to be tackled. In his calculations, MacKay does not assume energy demand continues to grow; like the IPCC does, he takes today's consumption as a baseline. By increasing efficiency and some other measures, we can reduce this consumption without radically altering our lifestyles (except perhaps shifting to public transport).

Both the IPCC and MacKay approach technology as a neutral instrument: we use it for our own goals, and those goals are not problematized. This is why vehicle fuel efficiency is discussed, but rarely the demand for transportation. We apparently have a certain need to move around, and fuel-efficient cars will make that happen with lower emissions than the current fleet of cars. What is left out of the question here is that demand for transportation is not fixed: we drive our cars because we have access to an excellent road network, because in our cities working and living happen in separate areas, and because driving appears to be inexpensive. Actor-network theory (Latour 2005) is useful for showing that how we got here has little to do with innate human needs, but everything with a contingent and messy history of industrialization of traffic. Once we see that, we notice that much more than the fuel efficiency of cars can change in the future.

The trust in efficiency improvement as a strategy for lowering consumption of resources in general is problematic: when increased efficiency lowers the price of a commodity, its demand tends to go up, resulting in a net *increase* in resource use. This effect was first noted in 1865 and is called *Jevons's Paradox* (Smil 2010). Economists speak of 'rebound effects' in a somewhat broader sense of reallocating resources for more of the same (direct rebound effects, keeping more lights on with the introduction of energy-saving light bulbs) and more of other things (indirect rebound effects, going on holiday with the money saved from the energy bill after insulating the house) (Sorrell 2007).

There are roughly two ways to go around this: to moralize the system, and to moralize the agents. The first route means that any savings that are made with efficiency improvements are taken out of the market, for example with taxes. If fuel prices increase synchronously with efficiency improvements, the costs of driving remain the same for the user, but the environmental burden declines. It is important that the taxes are spent in a way that does not create new environmental costs. The second route is to ask of people to let ethical rather than economic considerations prevail when contemplating consumption. Public service advertisements could make people question whether they could cut down on driving. In this case it is important that indirect rebound effects are covered as well. An example of such behavior can be found in *voluntary simplicity* (Elgin and Mitchell 1977): the choice to work less and live on a lower income (and therefore at a lower consumption level) in return for more time spent on intrinsically valuable activities.

Both these routes need the other. In order for environmental taxes to be acceptable to a society, the members of this society need to have some understanding of environmental concerns and how their behavior influences them; in order to make moral choices on consumption, society needs an infrastructure to make these choices: alternatives for transportation by automobile, a job market that offers meaningful part-time work, etc.

This co-dependence is further complicated by the active role that technologies may play in this dynamic. It has been argued that morality is embedded in technological artifacts and systems, and that they even function as moral agents (e.g., Winner 1986; Latour 1992; Verbeek 2005). If this position is accepted, the task is not just to create citizen awareness and environmentally sound rules of the game, but also to reinvent technologies in ways that make them for example invite awareness and inhibit wasteful behavior. Still, this cannot be the only focus; it is impossible to delegate *all* morality to technology, as long as it is people who make the policies and design new artifacts: we cannot expect an environment of moralized technology designed by a society that is ignorant of the issues at stake.

Philosophy of Technology-Informed Approaches to Climate Change

In most of the literature mentioned thus far, climate change is seen as a design challenge, a negative consequence of current practices, while those practices themselves are unproblematic. The goal is to do what we already do, but without these actions burdening the climate. In this section the scope of the problem is expanded. In one way or another, the authors reviewed argue that climate change-inducing activities are problematic *in themselves*, and that a certain critique of technology helps in clarifying the problem.

Footprint Minimization

In the wake of the seminal report *Limits to Growth* (Club of Rome 1972), an environmental design movement emerged (e.g., Schumacher (1973)). The emphasis was put on scale: whereas

these commentators noticed an ever-increasing scale to the practices around them, they turned this radically around by working on a scale as small as possible. This was supposed both to solve environmental problems and to increase the quality of life. Traces of these ideas can be found in more recent discourses on the circular economy (McDonough and Braungart 2002), autarkic building (Frenay 2009), and the degrowth movement (Demaria et al. 2013).

For these authors, environmental concerns gave rise to the idea that the way society organizes itself is wrong. Either because our practices do not agree with our environment or with our own human nature, the practices need to be turned around. Although they do not make it very explicit, these authors realize that our technologies and technological systems imply certain practices and behavior. It is this insight that leads them to come up with alternative designs for buildings, gardens, and industrial life cycles in order to make our lives better. This is where input from philosophy of technology becomes very valuable. Two thinkers in particular can provide us with useful concepts in this area: Ivan Illich and Albert Borgmann.

(Energy) Consumption and the Good Life

Neither Illich nor Borgmann were concerned with climate change, but rather with the problems of industrial society. However, because greenhouse gas emissions are strongly correlated with industrial production and consumption, critiques of industrial society can certainly help inform climate change mitigation efforts. Although their interests were similar, Illich and Borgmann approached their subject from diametrically opposed positions. Illich started at the level of society and focused on the workings of the industrial system at large to come up with a set of negative rules within which meaningful human life could still be possible; Borgmann started at the level of the individual and the effects of industrial society on them to come up with positive rules on what a meaningful life should involve. Together they offer a powerful vocabulary to rethink the connection between consumption and quality of life.

Ivan Illich wrote extensively about many facets of modern society, criticizing its blind faith in narrow goals from education and health care to transportation. *Tools for Conviviality* (1973) most broadly develops his theory; the short book *Energy and Equity* (1974) and the essay *The Social Construction of Energy* (2013) deal more specifically with energy practices and therefore with climate-related concerns.

The central concept in Illich's critique of industrial society is *overefficiency*: the tendency to focus on a specific goal, which becomes, after time, countereffective. An example is the speed of traffic. The automobile appears to offer speed and convenience to its user, because the typical speed at which we drive our cars is much higher than that of a cyclist or a pedestrian. But when we add up all the costs of the car (finding parking spots; working to pay for gasoline, road tax, monthly installments, insurance, etc.), it turns out the automobile is on average slower than a bicycle, and only a little faster than a pedestrian. Meanwhile, cities expand, and the need for transport increases, so we spend more rather than less time getting around:

What distinguishes the traffic in rich countries from the traffic in poor countries is not more mileage per hour of life-time for the majority, but more hours of compulsory consumption of high doses of energy, packaged and unequally distributed by the transportation industry. (Illich 1974: 19)

Motorized traffic has become a *radical monopoly*: it has created a need that only it can satisfy. This need costs time and resources, creates inequalities (the entire society pays for services only the elite can enjoy), and reduces the possibility of other ways of life (a city designed for automobiles is difficult to walk in).

The needs that industry creates are not human needs, according to Illich, and indeed they actually tend to get in the way of our real needs. In order to overcome this tendency, Illich argues that society needs to put boundaries on its energy use. Where exactly these boundaries lie is up for debate, but the main point is to stop them from expanding beyond the onset of over-efficiency. Illich describes good tools as those which

can be easily used, by anybody, as often or as seldom as desired, for the accomplishment of a purpose chosen by the user. The use of such tools by one person does not restrain another from using them equally. They do not require previous certification of the user. Their existence does not impose any obligation to use them. They allow the user to express his meaning in action.

(Illich 1973: 22)

This position may be utopian, but Illich at least makes a good case for being more critical of the kind and amount of consumption we understand as necessary or valuable. What this means for climate policy is that the discussion should go beyond who must take the greatest burden and which technologies can provide us with climate-neutral alternatives, but rather should raise the question what climate-burdening practices are merely fulfilling their own needs, while adding little or nothing to human quality of life.

Such questioning could further benefit from the work of Albert Borgmann, who wrote extensively about the relationship between technology and quality of life (Borgmann 1987). For Borgmann, the defining characteristic of modern technology is its tendency to commodify things of value: making them available instantly while hiding the machinery that brings forth these commodities. The resulting structure of technological societies he calls the *device paradigm*. In their transformation from things to commodities, something of value gets lost, according to Borgmann. A common example is the way we heat our homes: where people used to have a central fireplace that needed maintenance and the chopping of wood before winter, modern houses are heated by central heating systems, often tucked away in a closet and fed by an underground natural gas pipeline. A thermostat allows the inhabitants to easily adjust the heat, and in many cases to program the system so that it is already heated up by the time they return from work. At a superficial level this seems to be unproblematic, but the issue for Borgmann is that the fireplace used to be more than a source of heat. A fireplace was a focus, the center of the house, a place where family members gathered to interact with each other. Chopping wood was not just a way to make sure one did not freeze to death in the winter, but it also connected people with their surroundings, so they learned to appreciate the gifts of nature. By commodifying heat, these engaging practices were lost and a one-dimensional consumption remained. Disengaging technologies like a central heating system are called *devices*, as opposed to traditional things.

The device paradigm replaces *focal practices*—that is, activities that are valuable in themselves. Focal practices can be aided by (modern) technology, in the shape of *instruments*. Examples are running shoes and freeze-dried food that enable us to explore trails in the wilderness, musical instruments that allow us to make music with friends, or kitchen appliances that help us prepare our own meals. People do not engage with instruments, but through them: the engagement with the focal things (the trail, the musical piece, or the shared meal) is mediated by instruments. The example of the hearth may suggest a rather nostalgic understanding of the traditional situation, but Borgmann's position should not be understood as a plea to go back to the 'good old days.' The fact that the distinction is most easily explained by comparing a traditional and a modern practice does not mean that modern technology is always and intrinsically bad. This is also made clear by Borgmann's embrace of modern instruments. However, Borgmann does notice a *pattern*, a tendency towards the device paradigm. This pattern is not

enfeebled by a few counterexamples. After making his perspective plausible, Borgmann argues that we should develop a sensitivity to the kind of usage particular technologies tend to invite and inhibit, and that we should surround ourselves with instruments rather than devices.

The distinction between devices and instruments becomes interesting for climate change mitigation efforts when we realize that in many instances, commodification coincides with increased energy consumption. Take, for instance, the tendency to spend holidays farther and farther away from home. Distance is commodified by air travel, which radically increases energy expenditure, but which does not necessarily make for better vacations, as is suggested by the classic *adagium* that “the journey is the destination.” Instruments like hiking boots or bicycles are arguably much better for experiencing the journey for what it is than the sterile environment of a modern airliner. This connection between devices and energy is probably not an iron rule, but if we are to reduce energy consumption to become climate neutral, a close look at devices does make sense.

The Energy Perspective on Development of Society

An entirely different approach to understanding the problem of climate change can be found in energy-centric literature. In this discourse, the development of human society is defined by its ability to appropriate increasing amounts of energy. The neolithic (agricultural) and industrial revolutions were qualitative leaps in this development, which resulted in an increase in both carrying capacity of an area and average per capita wealth. Agriculture made food much more readily available; the industrial revolution greatly increased the availability of exosomatic energy, creating something of a technological metabolism. This is a continuation of the tendency of life to evolve towards increasing energy conversion (see, e.g., White 1959) and complexity. In an intuitive sense, one may imagine the development of microbes, photosynthetic microbes, plants, herbivores, carnivores, and fossil-fuel consuming humans as ever more complex ways to expend energy (Bataille 1988). Climate change problematizes this development, because it suggests that the appropriation of fossil energy is not sustainable. At the same time, this perspective on life problematizes the climate change mitigation strategy of reduced energy consumption, as this seems to go against our nature.

From here, a number of outcomes are imaginable. (1) *Boom-and-bust*: society fails to act on climate change, ecosystem services collapse, civilization collapses. Nonhuman life and the ecosystem at large may suffer as well. (2) *Continuous progress*: renewable alternatives are developed to continue the path up towards greater energy expenditure. This may be possible because the planetary solar income is still orders of magnitude larger than human energy usage, and some futurists expect humanity to expand to outer space as well. (3) *Respecting the niche*: society manages to constrain itself and only uses energy and other resources (land, water, etc.) at a lower, renewable pace. A relatively stable balance could be found.

What these three scenarios have in common is their abstraction from the details of the present situation, and the focus on the larger timeline of which we are part. Yes, we may now be in a position to steer life in this or that direction, but it is ultimately not about humanity but about the development of life, which, at the present, has humanity at its forefront. Such a perspective may inspire all kinds of moral positions (we are so small and insignificant; everything is permitted; there are greater forces at work; etc.), but the perspective should give us pause. Bataille described it as follows:

The beings that we are are not given once and for all; they appear designed for an increase of their energy resources. They generally make this increase, beyond mere subsistence, their goal and their reason for being. But with this subordination to increase, the being in

question loses its autonomy; it subordinates itself to what it will be in the future, owing to the increase of its resources.

(Bataille 1988: 190)

If we assume some kind of human autonomy, a sense that we can direct ourselves to what we *are*, rather than what we could *become*, it becomes clear that the life's tendency to increasing energy expenditure is not particularly beneficial for us humans. Similar to the herbivore who does not benefit from the development of the carnivore, the human does not necessarily benefit from the layer of exosomatic energy expenditure that is an important cause of climate change: it may lead exactly to the traps that Borgmann and Illich noticed in their critique of industrial society.

Climate Change Informing Philosophy of Technology

Beyond Artifacts

Recently, philosophy of technology has focused mostly on artifacts, human-technology interaction, and development of novel technologies (e.g., information technologies, nanotechnologies). Large systems of already commonplace technologies have remained unacknowledged in these contexts. It is largely these technological systems that are causing climate change: cities built around the automobile, global trade and travel, concrete buildings, artificial fertilizers, etc. Novel technologies may of course have a big impact on climate change adaptation and mitigation. 'Smart grids,' for example, use information technology to distribute and manage decentralized energy networks. But at the same time, they are based on the system already in place: on the semiconductor industry, on electricity as the universal form of energy. Some degrowth thinkers hold that it will prove impossible to keep that basis intact if it is to become truly sustainable, so they problematize any proposed solution that is built on top of it. How can we understand the connection of such systems? A *systems perspective* in philosophy of technology would be helpful to appreciate and assess such claims. To what extent must we understand technological developments as incremental layers on top of the sediment of older technologies, and to what extent can they truly depart from the system on which they were built? Such questions would not only be useful in the context of climate change and sustainable development, but also in the discourse on responsible innovation and disruptive technologies.

Technology in the Anthropocene

A systems perspective in philosophy of technology would not just align with the techno-social systems that cause climate change, but also with the understanding of the climate system itself, and its part in the even larger context of Earth System Science, the scientific niche concerned with the interplay between the atmosphere, geology, and the biosphere. Climate change is one way in which it becomes clear that human behavior now plays a key role in the development of the Earth. It is therefore part of a discourse on the *Anthropocene*, or the geological epoch in which humans are the dominant factor in the shaping of the Earth (Crutzen and Stoermer 2000). A lively debate has emerged on both the accuracy and implications of this concept. It is appreciated as a useful concept to emphasize connections between various global ecological concerns, as well as critiqued because of its alarmist and technophilic connotations (Baskin 2015). But whatever one's normative position on the matter, the Anthropocene, understood as the ultimate consequence of technology at the global scale, forces us to return to theories on technology at

that scale. Technologies do not just change the way we see (like glasses do) or what we consider friendship (like Facebook does); they change the very planet we live on.

Geoengineering is typically understood as the practice of deliberately influencing the climate through technologies such as albedo-increasing aerosols or radiation-blocking satellites. It is proposed as a risky but possibly life-saving backup plan if society fails to reduce its CO₂ emissions in time to avoid serious catastrophes (Blackstock et al. 2009). But this narrow understanding of geoengineering misses the fact that unknowingly we have been ‘engineering’ the climate for centuries. Climate change started as a side effect, a regretful result of industrial society, a design flaw. But now that climate science has given us the first imperfect understandings of how our behavior is changing the climate, any attempt to react to this in our behavior is a form of conscious geoengineering: every new coal-fired power plant is being built while we are aware of the amount of CO₂ it will emit in its lifespan, and the effects this will likely have on the planetary climate. If we are to do justice to the idea of the Anthropocene, any discussion on climate-influencing policy and technology should be approached as a discussion on geoengineering in this broader sense.

If in the end we manage to master the challenge of climate change and succeed to transition toward a sustainable society, we may find ourselves in a paradoxical situation related to the concept of Anthropocene. We are currently struggling to stay away from catastrophic temperature rises, but it is conceivable that at a certain point our climate managing strategies are effective enough that we may choose to restore the preindustrial atmospheric balance. For geologists, the boundaries of the Anthropocene epoch are defined by markers in sediment such as radioactive isotopes that resulted from atmospheric nuclear tests in the 1950s, and the steep increase of CO₂ concentrations from roughly the same period. But what happens if such markers disappear, not because we stop influencing the atmosphere but because we develop perfect control over it? Would that mark the end of the Anthropocene, or rather the true beginning of it?

Philosophical puzzles aside, this chapter makes clear that climate change and philosophy of technology challenge each other. To understand difficulties in climate change mitigation efforts, philosophy of technology provides some useful concepts and insights. At the same time, climate change and the perspective of the engineered global atmosphere provide new challenges for philosophy of technology. It is to be expected that climate issues become more pressing in the future, so it is to be hoped that the interface between the two gains more attention from philosophers in the years to come.

References

- Baskin, J. (2009). The Impossible Necessity of Climate Justice? *Melbourne Journal of International Law*, 10, 424–439.
- Baskin, J. (2015). Paradigm Dressed as Epoch: The Ideology of the Anthropocene. *Environmental Values*, 24(1), 9–29.
- Bataille, G. (1988). *The Accursed Share: An Essay on General Economy*. New York: Zone Books.
- Blackstock, J. J., Battisti, D. S., Caldeira, K., Eardley, D. M., Katz, J. I., Keith, D. W., . . . Koonin, S.E. et al. (2009). *Climate Engineering Responses to Climate Emergencies*. IOP Conference Series: Earth and Environmental Science. Santa Barbara, CA: IOP Publishing.
- Borgmann, A. (1987). *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago: University Of Chicago Press.
- Crutzen, P. J. and Stoermer, E. F. (2000). The Anthropocene. *Global Change Newsletter*, 41, 17–18.
- Demaria, F., Schneider, F., Sekulova, F. and Martinez-Alier, J. (2013). What Is Degrowth? From an Activist Slogan to a Social Movement. *Environmental Values*, 22(2), 191–215.
- Elgin, D. S. and Mitchell, A. (1977). Voluntary Simplicity: Life-Style of the Future? *The Futurist*, 11(4), 254–261.

- Frenay, M. (2009). Earthships: Sustainable Housing Alternative. *International Journal of Sustainable Design*, 1(2), 223–240.
- Gardiner, S. M. (2006). A Perfect Moral Storm: Climate Change, Intergenerational Ethics and the Problem of Moral Corruption. *Environmental Values*, 15, 397–413.
- Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162(5364), 1243–1248.
- Illich, I. (1973). *Tools for Conviviality*. London: Marion Boyars Publishers.
- Illich, I. (1974). *Energy and Equity*. New York: Harper and Row.
- Illich, I. (2013). The Social Construction of Technology. In Sajay Samuel (ed.) *Beyond Economics and Ecology: The Radical Thought of Ivan Illich*. London: Marion Boyars Publishers, 105–123.
- IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva: Intergovernmental Panel on Climate Change.
- Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. In W. Bijker and J. Law (eds.) *Shaping Technology*. Cambridge, MA: MIT Press, 225–258.
- Latour, B. (2005). *Reassembling the Social—an Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- MacKay, D. J. (2009). *Sustainable Energy—Without the Hot Air*. Cambridge: UIT Cambridge Ltd.
- McDonough, W., and Braungart, M. (2002). *Cradle to Cradle: Remaking the Way We Make Things First*. New York: North Point Press.
- Meadows, D. H., Meadows, D. L., Randers, J., and Behrens III, W. W. (1972). *The Limits to Growth, Club of Rome*. New York: Universe Books.
- Posner, E. A., and Weisbach, D. (2010). *Climate Change Justice*. Princeton, NJ: Princeton University Press.
- Schumacher, E. F. (1973). *Small is Beautiful: Economics as if People Mattered*. New York: Harper and Row.
- Smil, V. (2010). Science, Energy, Ethics, and Civilization. In R. Y. Chiao, M. L. Cohen, and A. J. Leggett. (eds.) *Visions of Discovery New Light on Physics, Cosmology, and Consciousness*. Cambridge: Cambridge University Press, 709–729.
- Sorrell, S. (2007). *The Rebound Effect: An Assessment of the Evidence for Economy-Wide Energy Savings From Improved Energy Efficiency*. Sussex: UK Energy Research Centre..
- Verbeek, P.-P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Penn State University Press.
- White, L. A. (1959). Energy and tools. In *The Evolution of Culture: The Development of Civilization to the Fall of Rome*, 33–49. New York: McGraw-Hill.
- Winner, L. (1986). Do Artifacts Have Politics? In *The Whale and the Reactor*. Chicago: University of Chicago Press, 19–39.

Outer Space and Imagination



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Space Tourism and Science Fiction

Rosalyn W. Berne

On February 25, 2015 the *Nightly Show With Larry Wilmore* (Stewart) featured a conversation between actor Don Cheadle, space-colonizer volunteer Sonia Van Meter, comedian Ricky Velez, and actor/film producer/director Lee Daniels. The subject of their discussion was Mars One, an organization based in the Netherlands that is seeking to land the first humans on Mars. Show host Larry Wilmore announced that the plan of the private enterprise is to take volunteers to the red planet and leave them there, in order to establish a permanent human colony. What seemed most concerning, or perhaps most curious, for the discussants on the show was Van Meter's willingness to leave her husband and children behind to take the one-way trip.

As a new business enterprise, space tourism (aka 'personal spaceflight' and 'citizen space exploration') has captured the attention of elected representatives, pundits, and the media. Opinions vary as to its economic viability, its technical feasibility, the candidate selection process, and the health and safety risks involved. As subject matter for the philosophy of science and technology, space tourism elicits a range of socioethical considerations, some of which have been reified by science fiction (SF).

When Wilmore asked his guests their opinions about the moral implications of a private company sending people to their certain death, candidate Van Meter replied, "The point is not to go to Mars to die, but to go out there to live!" (Stewart 2015). As is the nature of the *Nightly Show*, this serious subject was broached through satire. For example, when Wilmore asked if diversity should be a factor of candidate selection Cheadle chimed in humorously, "No; let all the White people go!" According to the Mars One website, there were 202,586 initial applicants. This would indicate strong public interest in traveling beyond the Earth's atmosphere, even if it means not coming back.

On the website of XCOR, another company offering future trips into space, it states, "These days, space is no longer the exclusive domain of governments and institutions like NASA and ESA. It is rapidly coming within reach of private organizations and people" (XCOR 2015). XCOR ticket prices are about \$100,000 per flight. But for some, access to an XCOR flight won't be a matter of financial ability: it can be won. Space Expedition Corporation, which manages XCOR trips, announced in December 2013 twenty-three winners of a yearlong competition for tickets to fly as soon as 2014 or 2015. (No flight had been taken as of the date of this writing as the space plane was still under development.) Astronaut Erwin "Buzz" Aldrin was at the Kennedy Space Center for the award ceremony. In his speech he said to the winners, "With my generation, only the few could go up into the heavens. When I took my first step on the moon I never could have guessed that space travel would evolve the way it has" (Aldrin 2013: n.p.). As a science fiction (SF) author, Aldrin probably *could* have guessed: His novel *Encounter With Tiber* (Aldrin 2005), co-authored with SF writer John Barnes, addressed the lack of funding for the space programs, space tourism included. The book begins with a near-future

history of how space becomes accessible to the general public. After its depiction of a second space shuttle accident, the problem is posed of how money is to be raised to pay for a new generation of fully reusable space vehicles.

Science Fiction (SF) as a Reaction to Social and Technological Change

Scientist and SF author David Brin's *Uplift Universe* collection (2013) depicts a galactic civilization that "uplifts all forms of life" with star-travelling ability. Humans are the only one of those capable forms of life, though in Brin's world, by comparison, humans are the weakest of all life forms. Regarding SF, Brin (Brin 2003) writes:

Many people have tried to define science fiction. I like to call it the literature of exploration and change. While other genres obsess upon so-called eternal verities, science fiction deals with the possibility that our children may have different problems. They may, indeed, be different than we have been.

(186)

The capacity to leave Earth, and to travel beyond its atmosphere as tourists, explorers, and possibly settlers, reflects the expansiveness of human knowledge and ambition. As our curiosity deepens, our imagination magnifies, and our scientific knowledge and tools become increasingly sophisticated, we have to come to terms with the inevitability of change. SF can play an important role in our doing so.

Functioning as rhetorical devices, the symbolic imagery, compelling dialogue, and imaginative, visualized extrapolations employed by SF can evoke meaning making beyond familiar domains of present reality. In this way, SF can be persuasive, enticing film viewers and fiction readers to contemplate alternate perspectives on what life may be, where life may exist, and who we as human beings may be when no longer bound to the Earth. And because SF also portrays spaces, technologies, and experiences that humans have not yet, but may eventually have access to—for better or for worse—it can summon ethical considerations. Isaac Asimov wrote, "Science fiction can be defined as that branch of literature which deals with the reaction of human beings to changes in science and technology" (1975: 92). In which case, SF portrayals of space tourism could be a reaction to new innovations in fuel sources and propulsion capacity; realizations that Earth's resources are being depleted; new understandings in astronomy, physics, and astrobiology; or evolving perspectives from the Earth Sciences suggestive that the planet may one day become uninhabitable. SF about space tourism may also simply be a reaction to the fact that new technological developments are making it possible for people to travel into space. That alone represents a significant change to what was once understood to be possible.

Going back into history, philosopher George Tucker published *A Voyage to the Moon* (Tucker 1975) in 1827, in the midst of the industrial revolution. People were still using horses and buggies, bikes, and walking for transportation, but steam-powered ships had crossed the Atlantic, and steam railways were coming on line: "By the end of the 19th century, international transportation undertook a new growth phase, especially with improvements in engine propulsion technology of the steamship and a gradual shift from coal to oil in the 1870's" (Rodrigue 2013: n.p.). Tucker's science fiction was most likely a reaction to the industrial revolution of his times. It extrapolates the new engines under development as so remarkably powerful, that they might eventually be able to transport humans to the moon! His writing also ridiculed the

social manners, religion, and professions of his colleagues, criticizing what he considered to be erroneous scientific methods of his day. As with most SF, Tucker's story was as much about the author's world, in the midst of radical scientific and technological change, as it was about an imagined world of the future.

By the time Jules Verne published *From the Earth to the Moon* (French, *De la terre à la lune*, 1865), invention of the electric motor was under way, and with new knowledge of electromagnetism the advent of electricity was about to begin. *From the Earth to the Moon* works through some of the physical and economic obstacles to success in landing a person on the moon. Indeed, a launch to the moon does happen in the novel, but it's not until the book's sequel, *Circling the Moon* (Verne and Walter 2010) that the destiny of Verne's fictional astronauts is revealed. As it turns out, they did enter lunar orbit, but rather than landing as planned, they are plunged into the cold, dark side of the moon. Eventually they emerge back into the light where they make additional attempts to land on the moon. One could interpret its meaning as reflective of humanity's movement into and back out of the dark ages, with the advent of the industrial revolution being our salvation.

It was another hundred years before reality caught up with science fiction. Apollo I was the first planned, manned mission of the U.S. Apollo manned lunar landing program, with a test target launch date of February 21, 1967. But in January of 1967, during a preflight test, a cabin fire killed all three crewmembers and destroyed the command module (Ertel I, Newkirk, R. et al. 1967). In Verne's novel, though the initial effort to land on the moon failed, there were no fatalities: its fictional threesome was projected from the moon's orbit back to Earth, where they fell into the sea, survived, and were celebrated as the first humans to leave the Earth.

In *The Fountains of Paradise* (Clark 1979), science-fiction writer Arthur C. Clarke introduced the idea of space elevators, a concept currently under design by engineering companies. Space rockets are very expensive to fuel, and a dangerous form of transport for passengers. A space elevator, in concept at least, would cost significantly less and carry fewer safety hazards. As an ABC news correspondent wrote, "It would also be a boon for space tourism" (referring to Obayashi, one of Japan's largest construction companies, which is working on a space elevator with cars that will carry thirty people up), "so it may not be too long before the Moon is the next must-see tourist destination" (Carney 2014). Obayashi aims to build its space elevator by the year 2050, "composed of a 96,000-km carbon nanotube cable, a 400-m diameter floating Earth Port and a 12,500-ton counter-weight" (Obayashi, n.d.). Historian Rodrigue suggests that such revolutionary changes in transportation tend to be rare but profound, "since they commonly involve the setting of entirely new networks" (n.p.). He notes further that while such innovations are not predictable, "once they occur it is possible to assess their potential impacts although the potential of an innovation can be exaggerated" (n.p.).

A discussion in the philosophy of science concerns representations of truth, and how scientific constructs might be compared to fictional ones. The history of science is replete with an inventory of concepts abandoned in the course of changes in knowledge; philosophers have reason to be skeptical. Barwich (2013) explains that "scientific discourse is permeated by idealized or often figurative descriptions and the issue is how literally to take them" (362). Science fiction depictions of space tourism don't have to prove themselves because:

Fiction deals with entities and descriptions that are not bound to be truthful descriptions of our world. Even though fiction contains entities that have familiar counterparts in the world, these elements are not automatically seen to serve as a truthful description of their counterparts.

(Barwich 2013: 362)

SF has no epistemic functions. But when, for example, Obayashi representatives say they are designing a space elevator that will “soar some 60,000 miles into the sky and feature magnetically powered cars that will carry people and cargo to a yet-to-be-built space station,” the issue at stake is the “epistemic function served by non-fictional elements” (Obayashi n.d., 362). Though the company does acknowledge that “current technology levels are not yet sufficient to realize the concept” (Carney 2014), the philosophical question becomes how to determine what makes for adequate epistemic use of such representations (i.e., whether a truthful claim is being made about the world). Barwich (2013) suggests using “proper fiction” in comparison with apparently fictional elements in science as a basis “on which the grounds for distinguishing between apparently fictional and non-fictional elements in science can be clarified” (363).

The Utility of “Proper Fiction”

When the boundary of truth and fiction is blurred, then the epistemological question is raised: how can it be known what actually is? Orson Welles broadcast H. G. Wells’ *War of the Worlds* in 1938, uncannily crossing the boundary between SF and known reality. Listeners panicked, afraid for their lives and for our planet. Meaning making at the boundary of SF and reality, between true and fictional representations, serves an important purpose in navigating scientific and technological change. Through the medium of SF, imagined worlds and other modes of being might be explored without threat of actual physical or psychological harm. *War of the Worlds* was aired at a time when looking up at the night sky evoked wonder and awe, but also trepidation: World War II was under way. Japanese planes had sunk American warships, the German military had been mobilized, and Hitler’s invasions had begun, taxing people’s sense of security. It’s no wonder that *War of the Worlds* caused quite a stir.

Technology changes what we know to be real. Fiction changes reality too, by redescribing it. Philosopher Paul Ricoeur (Ricoeur 1991) wrote:

The ultimate role of the image is not only to diffuse meaning across diverse sensorial fields, to hallucinate thought in some way, but on the contrary to effect a sort of Epoché of the real, to suspend our attention to the real, to place us in a state of non-engagement with regard to perception or action, in short, to suspend meaning in the neutralized atmosphere to which one could give the name of the dimension of fiction. In this state of non-engagement we try new ideas, new values, new ways of being-in-the-world. Imagination is this free play of possibilities. In this state, fiction can, as we said above, create a re-description of reality. (134)

The film *2001: A Space Odyssey* (Kubrick 1968) is set on a space station orbiting Earth, where people are awaiting transfer to Clavius base, a US outpost on the moon. The horror element of the film happens on *Discovery One*, a spacecraft bound for Jupiter, when the autonomous and increasingly maleficent ship’s computer, HAL 9000, turns off the life support functions of the crew. The meaning of the film is widely interpreted, but according to an independent film critic, Kubrick once said: “If the film stirs the emotions and penetrates the subconscious of the viewer, if it stimulates, however inchoately, his mythological and religious yearnings and impulses, then it has succeeded” (Ager, 2008). Only one year after the film’s release, on July 20, 1969, futuristic space travel was science fiction no more: *Apollo II* landed on the moon with Neil Armstrong and Buzz Aldrin. On their return to Earth from the moon, the two astronauts were celebrated as heroes. But because of the epistemological quandary of how to determine what is true of science and technology, there are some people today who remain skeptical of whether the landing on the moon actually happened (Launius, 2009).

The Economics of Leaving Earth

Potential financial gains are a significant factor of motivation in space tourism programs. The potential for new economic markets points to expansion opportunities for investors, and opens a new source of revenue for funding nationalistic and scientific endeavors in space. Russia, for one, has been able to continue funding its space program, in part, because of earnings from space tourism.

When addressing the U.S. Congressional Subcommittee on Space and Aeronautics, Edward L. Hudgins (Hudgins 2001) of the Cato Institute said,

The ability to travel outside the Earth's atmosphere already has ushered in important commercial benefits, most notably from communications satellites and remote sensing, the sectors in which private suppliers have been freest to operate. Data from the FAA's Office of Commercial Space Transportation places the value of America's space-related economic activity at 61.3 billion annually, generating earning of \$16.4 billion and employing nearly 500,00 workers.

(10)

That said, finances have been as much of an impediment as a motivation, in both fictional and nonfictional considerations of space travel. Aldrin spoke before that same subcommittee hearing. When Committee Chairman Dana Rohrabacher (California) introduced Aldrin (Rohrabacher 2001), he said,

Not many people know that you have got a Doctorate from MIT. I want to brag about you a little bit and let everybody know that we are very proud of you Buzz. . . . You are a famous man. Second man on the Moon and, more than that, you are a very patriotic man, and citizen who is deeply involved in space today. And trying to help guide our policies, especially on the issue of space tourism and opening up space for the average America.

(28)

Aldrin's testimony followed with the claim:

After 40 years of space exploration, space tourism has emerged as the key to generating the high-volume traffic that will bring down launch costs. NASA's own research has suggested that tens of millions of U.S. citizens want to travel in space, with far more if the global market is addressed. This immense volume of ticket-buying passengers can be the solution to the problem of high space costs that plague government and private space efforts alike.

(29)

One such ticket-buying passenger was Denis A. Tito, CEO of Wilshire Associates and the first paying civilian passenger to travel to the International Space Station (ISS). He, too, addressed the congressional subcommittee (Tito 2001). Tito told the committee that he had been interested in space since age 17, a passion ignited by the launch of Sputnik, the first artificial Earth satellite, and a Soviet Union accomplishment that came as a huge surprise to the United States and triggered the Space Race as part of the ongoing Cold War. And so it's ironic that Tito's ride to the ISS was on a Russian craft, an adventure he refers to as the greatest experience of his life. The Russian government was receptive to flying civilians on their spacecraft—for a fee.

At the time that Tito took that flight, the U.S. government was not amenable to civilian's flying, a hesitance that followed the tragic deaths of civilians on the Challenger flight of 1986,

including Sharon Christa Corrigan McAuliffe, who'd been selected from more than 11,000 applicants to participate as the first teacher in space. Tito touted the ideal of

having people from all walks of life, teachers, journalists, novelists, opera composers, etcetera, take their experience that they might have and bring it back to us on earth. So as a culture, we can better understand and experience the pleasure of flying in space and make that part of our culture.

(18)

Since Dennis Tito's flight to the ISS in 2001 (at a speculated personal cost to him of \$20 million) seven other people have spent upwards of \$40 million each for the same privilege, including Cirque du Soleil performer Guy Laliberté. Each flew as a 'spacecraft tourist' on the Russian craft Soyuz, developed by Energia, the largest company of the Russian space industry and lead developer of the Russian part of the ISS. That Russian space tourism program was suspended in 2009, but an announcement made in March 2015 suggests that the Russian Federal Space Agency is making plans for space tourist flights to the ISS as of 2018. Being largely a financial motivation, it appears to be that it is seeking to make up for the loss of a contract with NASA for the delivery of astronauts to the ISS.

Other Motivations

For Tito, the first civilian to fly into outer space, the experience was sublime. As he described it, "There was one thing not even the most extensive training could prepare me for; the awe and wonder I felt at seeing our beautiful Earth, the fragile atmosphere at the horizon, and the vast blackness of space against which it was set" (20). That reaction may have been the 'overview effect' reported by many astronauts on seeing firsthand the reality of the Earth in space. The term designates a shift in perspective, where national boundaries are removed and humanity melds into one planetary society, raising the imperative to protect the 'pale blue dot' that is our current home.

This is a sublime experience for sure, but given the safety risks and the extreme financial costs, why ought such an experience become a new cultural norm? For Tito the answer lies in the long-term value to society for expanding human civilization beyond Earth's orbit, and to ultimately be able to live in space—not just as a platform for scientific research, but for the capacity of civilians to live for long periods of time in space, a long-term vision that was once limited to the realm of SF.

In his *Religion of Technology*, David Noble (1997) claims that mythological yearnings are at the root of such ambitions to leave Earth. Stadler (1998) notes that for Noble, at its core, space technology embodies a tenet of religious millenarianism promising the transcendence of mortal life. As Stadler writes in his review of Noble's book,

Shooting people into space is read as the most literal attempt to leave Earth behind: to enter paradise physically. As the Apollo 11, the first manned capsule, landed on the moon in a spot called the Sea of Tranquility, Erwin Aldrin—Presbyterian, Sunday school teacher, and the second man on board (the other was Neil Armstrong)—asked Mission Control for radio silence. He then unpacked a small kit provided by his pastor, took communion, and read from the bible.

(4)

Noble's admonition is that such aspirations disregard the remedy of humankind's most pressing needs. In other words, what's the point of heading for space if it doesn't help improve the quality of life on Earth? That same question might be asked of any endeavor used solely to entertain,

or to fulfill a purely personal need for adventure. But what if the long-term intention on living off of Earth is indeed intended to address some of humanity's most pressing needs? There are those who claim that human existence on Earth may soon be threatened by devastating arrival of an asteroid, or that we may run out of the resources needed to continue to live here, such as water and energy.

The SF film *Interstellar* (2014) depicts one such scenario, wherein due to climate conditions a second and massive Dust Bowl has rendered Earth uninhabitable for human life. Dust has replaced fertile soil and crop diversity has been lost, genetically modified corn being the only crop still producing food adequate for human consumption. In the world of the film, human life is on the brink of extinction due to massive hunger. A secret plan has been undertaken to send a small cadre of humans into space in order to begin a colony, in a desperate effort to continue the human species. The premise of this fictional depiction of civilian space travel is that human life is so valuable that it must continue, even if off of Earth. (Perpetuation of the species at all costs.) The end of the story suggests that if we put our knowledge to technological aims then humanity will endure, albeit probably not on Earth.

To the extent that the SF film *Interstellar* is as much a reaction to current changes faced by humanity, as it is about some imagined future elsewhere in the universe, its most important function may be in its exploring the existential questions of who we are as human beings, what our purpose is, and the capacity of individuals as well as collective humanity to open to possibilities of living life beyond its origins on the planet Earth. In the film, the character of Cooper (played by Matthew McConaughey) says,

We've always defined ourselves by the ability to overcome the impossible. And we count these moments. These moments when we dare to aim higher, to break barriers, to reach for the stars, to make the unknown known. We count these moments as our proudest achievements. But we lost all that. Or perhaps we've just forgotten that we are still pioneers. And we've barely begun. And that our greatest accomplishments cannot be behind us, because our destiny lies above us.

When SF and Space Tourism Meld

The SF film *Fifth Element* (1997) takes place inside of an orbiting hotel with features much like an ocean cruise liner. Though the film has its critics, its imaginary expeditions are not that far off from the stated ambitions of today's private space tourism endeavors. The 'dream' of the Japanese Shimizu Corporation envisions a proposed Space Hotel where participants will "spend pleasant time viewing down the transparent blue Earth, thin veil atmosphere, beautiful floating clouds and the dawn of the Earth. In addition, they will enjoy astronomical observation, sports, and meals under microgravity as well as communications with the Earth" (Shimizu n.d.). A U.S. commercial spaceflight company, Virgin Galactic is aiming to create the world's first commercial space line. The Virgin Galactic website reads,

The roughly 700 Virgin Galactic future astronauts who have already paid deposits for their flights on SpaceShipTwo come from more than 50 different countries, about half of which have never before sent a human to space. They span in age from under 10 to over 90 years of age. They practice many professions and speak many languages.

(Virgin Galactic 2015)

As explained, the inspiration is "to inspire future generations and make it possible to see the beautiful planet we call home from a new perspective" (Virgin Galactic 2015). To the company's

thinking, such a new perspective will bring benefits to humans living on Earth, claiming that “in the future, life on Earth will be made better by the exploration of space” (Virgin Galactic 2015) by inspiring children who will eventually start new businesses and companies, by providing experiences that will help launch leadership careers, and by delivering data that may help save lives and identify new resources. Perhaps this kind of space tourism, with its broad, far-reaching socioeconomic purposes, is not mythological or religious in nature as Noble might suggest, as it seems to be refocusing attention back on Earth rather than to motivate leaving it.

Another way of understanding the philosophical significance and meaning of space tourism, both in its fictional representations and nonfictional, practical endeavors, is as a mythic journey of frontier tourism. J. Laing and G. Crouch (2011) characterize frontier tourism as visiting places that “lack a permanent resident population,” which “requires a tremendous level of preplanning, including, in many cases, fundraising, logistics, communication and training” (1517). They write,

Psychological preparation is also required for the difficult circumstances that may be faced by the frontier tourist, including loneliness, sensory deprivation, limited rations, harsh climatic conditions and high levels of danger, even where a guide is present and it is possible to communicate with the outside world.

(2011: 1517)

Seeing the division between tourism and exploration blurring, Laing and Crouch have sought to understand what might be fueling a push towards the frontier. They posit a number of possibilities, including a postmodern fixation with novelty and myth making, which takes us into a “third age of exploration.” They’ve researched individuals who have engaged in frontier tourism, finding that some individuals “re-enact archetypal myths based on the epic tales of exploration and, in the process, contribute to the modern mythology of the heroic journey for a new generation of tourists” and in doing so, “feed into modern mythologies, potentially inspiring a new generation to seek the unknown and test themselves in frontier environments” (1530). Valentine, Olsson, and Battaglia (2012) present evidence for

how people deal with the demands of becoming-in-the-cosmos by using extremes as a limit/horizon relation that points to an unknown what-may-be, while simultaneously finding ways of naming it. It follows that both limit and horizon are in play at all times in extreme sites, even as they may seem—or may be represented as—opposed.

(1008)

And so it is with SF and space tourism: it comes functioning in a dialectic of limit and horizon, the visualized and named, balanced between the fictional imaginings of those who use dialogue, setting, plot and imagery to navigate change, and the laid plans of those who aim to use technology to take individuals beyond the confines of Earth. Valentine et al. (2012) “push for us to think about the extreme not only as a trope of human embodied experience and not only in a technoscientific context, but as a *cosmological* explanation of humanness that exceeds modernity’s everyday spaces and timelines” (1012). Humans are inclined to push against boundaries, both physical and mythological, to venture into frontiers, both imagined and real, to make meaning of what it means to be human, and to seek to expand beyond the apparent limits of our Earthly selves.

References

Ager, R. (2008). *Kubrick and Beyond the Cinema Frame: An In-Depth Analysis of 2001: A Space Odyssey*. Available at www.collativelearning.com/2001%20chapter%201.html [Accessed March 27, 2008].

- Aldrin, B. (2013). 23 Axe Apollo Fans With the Right Stuff Win Free Space Trips. Available at: <https://buzzaldrin.com/23-axe-apollo-fans-with-the-right-stuff-win-free-space-trips/> [Accessed April 5, 2015].
- Aldrin, B., and Barnes, J. (2005). *Encounter With Tiber*. New York: Open Road.
- Asimov, I. (1975). Review: How Easy to See the Future! *Natural History*, LXXXIV(April), 92.
- Barwich, A. (2013). Science and Fiction: Analysing the Concept of Fiction in Science and its Limits. *Journal for General Philosophy of Science*, 44(2), 357–373. doi:10.1007/s10838-013-9228-2
- Brin, D. (2003). Tomorrow May Be Different. In Karen Haber (ed.) *Exploring the Matrix: Visions of the Cyber Present*. New York: MacMillan.
- Carney, M. (2014). *Japanese Construction Giant Obayashi Announces Plans to Have a Space Elevator Up and Running By 2050*. Available at: www.abc.net.au/news/2014-09-21/japanese-construction-giants-promise-space-elevator-by-2050/5756206 [Accessed March 30, 2015].
- Clarke, A. C. (1979). *The Fountains of Paradise*, 1st edition. New York: Harcourt Brace Jovanovich.
- Ertel, I., Newkirk, R., et al. (1967). *Part 1 (H): Preparation for Flight, the Accident, and Investigation: March 16 Through April 5, 1967*. Available at: www.hq.nasa.gov/pao/History/SP-4009/v4p1h.htm [Accessed March 31, 2015].
- Hudgins, E. L. (2001). *Space Tourism: Opening Statement*. Washington, DC: U.S. G.P.O..
- Kubrick, S. (1968). *2001: A Space Odyssey*.
- Laing, J. H. and Crouch, G. I. (2011). Frontier Tourism: Retracing Mythic Journeys. *Annals of Tourism Research*, 38(4), 1516–1534.
- Noble, D. F. (1997). *The Religion of Technology: The Divinity of Man and the Spirit of Invention*, 1st edition. New York: Alfred A. Knopf.
- Obayashi. (n.d.). *The Space Elevator Construction Concept*. Available at: www.obayashi.co.jp/english/special/2014110424.html [Accessed March 31, 2015].
- Orwell, G. (1938). *War of the Worlds*. The Mercury Theater on Air.
- Ricoeur, P. (1991). The Function of Fiction in Shaping Reality. In V. J. Mario (ed.) *A Ricoeur Reader*. Toronto: University of Toronto Press, 117.
- Rodrigue, J. P. (2013). *The Geography of Transport Systems*. New York: Routledge. Segments available at: <https://people.hofstra.edu/geotrans/eng/ch2en/conc2en/ch2c5en.html>
- Rohrabacher, D. (2001). *Space Tourism: Opening Statement*. Washington, DC: U.S. G.P.O..
- Shimizu. (n.d.). Space Hotel—Space Tourism. Available at: www.shimz.co.jp/english/theme/dream/spacehotel.html [Accessed April 6, 2015]
- Stadler, F. (1998). The Religion of Technology: The Divinity of Man and the Spirit of Invention. *Canadian Journal of Communication*, 23(4), 25.
- Stewart, J. (Executive Producer) (2015). *2015 The Nightly Show with Larry Wilmore*, Television broadcast, February 25. Available at: www.cc.com/episodes/t3k3gt/the-nightly-show-february-25-2015-mars-2024-season-1-ep-01018
- Tito, D. (2001). *Space Tourism: A Statement*. Washington, DC: U.S. G.P.O..
- Tucker, G. (1975 [1827]). *A Voyage to the Moon*. Boston, MA: Gregg Press.
- Valentine, D., Olson, V. A., and Battaglia, D. (2012). Extreme: Limits and Horizons in the Once and Future Cosmos. *Anthropological Quarterly*, 85(4), 1007–1026.
- Verne, J. (1874). *From the Earth to the Moon Direct in Ninety-Seven Hours and Twenty Minutes, and a Trip Round It*. New York: Scribner Armstrong.
- Verne, J., and Walter, F. P. (2010). Circling the Moon. In F. P. Walter (ed. and trans.) *The Amazing Journeys [Electronic Resource]: Five Visionary Classics*, Excelsior editions. Albany: State University of New York Press, 223.
- Virgin Galactic (2015). Available at: www.virinalgalactic.com/ [Accessed March 31, 2015].
- XCOR website (2015). Available at: <http://xcor.com/flytospace/>

Well-Ordered Engineering

Participatory Technology Assessment at NASA

Zachary Pirtle¹ and David C. Tomblin

Introduction

How should democracies engage with complex engineering and science decisions? Many philosophers have advocated for a more just and fruitful connection between science, engineering and society. Philosopher Philip Kitcher (2001) created a conceptual model—well-ordered science (WOS)—for how science and engineering should decide on their goals. However, while several models of the relationship between engineering and democracy exist, very few have successfully utilized these models in practice, leaving the practical connection between science and engineering unchanged. Indeed, Kitcher argues that we do not have sufficient social science techniques to enable democracy to guide engineering and science. One key challenge is how to understand what everyday citizens would want science to achieve if they were sufficiently informed, in an unbiased way, for their views to become salient in decision-making contexts for engineers and scientists.

There has long been a strong foundation in social science that can greatly aid philosophers who want to help align engineering with what Kitcher (2001) considers a “society well-ordered.” This is a society where decisions about the three basic phases scientific research and technological development (agenda-setting, investigations, and applications) consider “the judgements of ideal deliberators, representative of the distribution of viewpoints in society” (123). We will illustrate the utility of social science by focusing on a single method that helps raise novel questions and issues that can accomplish part of what Kitcher proposed: participatory technology assessment (Scolve 2010). In particular, we focus on the participatory technology assessment (pTA) of NASA’s Asteroid Initiative. In this forum, citizens learned about technical decisions associated with NASA’s future plans, deliberated in a carefully structured manner, and then provided input on what NASA’s goals and decisions ought to be. NASA sought and valued this assessment as it sought to best align its work with what society needs.

In all, this forum emphasized a direction about how to ‘fill in the gap’ between philosophical literature like Kitcher’s well-ordered science and actual practical experience. Actually engaging citizens on real decisions creates a mix of complex data that both philosophers and reflective engineers can use to understand epistemological and practical aspects of their respective fields. NASA’s pTA example shows how social science can enrich philosophical study of science policy goals. While participatory technology assessments have occurred in the past, especially in Europe and to a limited extent in the United States, this example may be the first time public input has been used for a major engineering decision at a national level. The success of this effort may augur well for future practical efforts to steer engineering and science toward a democratically aligned goal.

Objective Knowledge and Democratic Input on Science Policy

Philosopher of science Philip Kitcher's book *Science, Truth and Democracy* is perhaps the most prominent attempt to align the philosophy of science and technology to discussions of science policy and democracy. The first half of the book contains a discussion of traditional issues within the philosophy of science that are structured to justify a more democratic approach to science policy. The second half of the book proposes an ideal for science policy, well-ordered science, that tries to articulate what science policy should look like. His 2011 follow-up book, *Science in a Democratic Society*, builds upon his vision for well-ordered science but does not fundamentally change his original framework. While we are focusing on Kitcher here for illustrative purposes, we must note that other philosophers engaged this topic, some much earlier than Kitcher. The work of Helen Longino (1990), Heather Douglas (2009), and Miriam Solomon (1994) should be part of a fuller discussion of how philosophy should think through the role of science in a democracy. There is additional reason to focus on Kitcher, as his theoretical framework is in some ways embodied in the pTA experiment discussed later.

Before we describe WOS, we will briefly sketch Kitcher's epistemological argument that motivates WOS. Kitcher describes the notion that the scientific pursuit of truth should be preserved against moral and ethical concerns external to science, which would imply that democratic decision-making processes need not directly be connected to science and engineering. The notion of holding science independent from society is driven by a belief that science is a morally virtuous project that is seeking to gain vitally important truths. If performing science automatically creates intrinsic and indisputable value for society, then there could be a moral imperative to free science and engineering from the broader society. However, Kitcher argues against the notion that scientists are pursuing some objectively important truths with a value that cannot be disputed (2001: 55–83). Despite many attempts, a universal, context-independent conception of scientific merit is unattainable, and Kitcher shows that scientific significance is constructed from a social origin. There can be no reliance on an overriding conception of the pursuit of truth to justify all scientific research.

The social nature of scientific significance has several implications. It suggests that research should be examined in terms of *both* its potential practical and epistemic benefits, with input about what science to fund from the broader lay public becoming highly valued. Democratic input on what science and engineering to fund becomes appropriate. Beyond the question of whether particular research is moral, the cost of science must be evaluated against opportunities lost: should society fund basic science, or pursue more direct social goods, or both? Why fund research on a multimillion dollar superconducting supercollider when millions of people live without potable water? Both basic and applied scientific research can generate good outcomes for society, but the risks of failure or of harmful repercussions should be evaluated by a democratic process. Based on this argument, Kitcher states that the scientific and engineering communities should align their work to develop knowledge that is more significant as determined by a democratic process. As political scientist Mark Brown shows, there are other alternative arguments for why a democratic approach to governing science and engineering is justifiable (Brown 2004). One can simply observe the effects that science and engineering, much of which is publicly funded, have on the world today to motivate governance and direction of science. As science and engineering greatly affect our lives, it is appropriate to put them within the system of checks and balances of the broader democracy. As Brown notes, Kitcher's arguments can have a particular strength because they place the impetus for democracy within the nature of science itself.

Well-Ordered Science and the Connection to Social Science

Given the lack of objective scientific guidance for determining broader research agendas, Kitcher asserts that democracy is the best mechanism to decide what science should be done. Given a commitment to democracy, Kitcher posits the ideal of well-ordered science, which would have representatives from all parts of society deliberating to determine what research is worth funding and how its benefits should be applied within society (see Table 29.1).

For a given issue, WOS serves as a thought experiment to help science policy participants imagine what a society would want its research to accomplish. Imagine convening the members of society, teaching them, and then providing them the help via arbitration to decide what science and engineering policies to implement. In a form that Kitcher attributes to John Rawls (Pirtle 2007: 65), he assumes that citizens overcome their deeply entrenched biases, such as political or religious orientation, and come prepared to make a consensus decision. The result would be a description of what research should be funded, how it should be conducted, and who should benefit. Kitcher hopes that scientists reflect on whether their work leads to goals that their society, if it had well-ordered science, would approve of. If scientists aspire to well-ordered science, they may be compelled to change their research, for instance to less controversial or more beneficial work. Kitcher's ideal also applies to the means by which science is done: if an idealized democratic evaluation might hold that a particular research approach is immoral, then a concerned investigator should consider whether or not to engage in that research.

Kitcher indicates that his philosophical ideal could not be fully implemented in all science policy endeavors.

There's no thought that well-ordered science must *actually institute* the complicated discussions I've envisaged. The thought is that, however inquiry proceeds, we want it to match the outcomes those complex procedures would achieve at the points I've indicated. Quite probably, setting up a vast populationwide discussion that mimicked the ideal procedure would be an extraordinarily bad idea, precisely because transactions among nonideal agents are both imperfect and costly. So the challenge is to find institutions that generate roughly the right results.

(2001: 123)

Effectively, Kitcher thinks the level of 'pulling together' of informed people with neutral viewpoints that engage and make serious recommendations about different decisions is too difficult. Kitcher wants social scientists to figure out ways to approximate the effects of the well-ordered science ideal, effectively filling the gap between scientific outcomes and public desire.

Table 29.1 Kitcher's Well-Ordered Science (WOS)

Imagine that representatives of society come together:

- with tutoring necessary to understand science policy decisions
- with a goal to understand and respect the preferences of other representatives
- with a commitment to making a decision by voting

In WOS, this group would vote and make science policy decisions at three levels:

1. The arbiters would decide on what the goals for scientific research should be.
 2. With the help of experts, the arbiters would approve of means to pursue and achieve these goals.
 3. Once research yields results, the arbiters would decide how to spread the benefits amongst society.
-

Was Kitcher right to have a division of labor between philosophy and social science? Brown argues there was already a foundation of good social science that Kitcher could have relied on. He rejects Kitcher's reliance on an *ideal* to try to get "roughly the right results" that match WOS. Brown argues that such reliance on ideals has throughout history been generally unsuccessful at actually generating reform, and that reliance on ideals can often lead to support of outcomes that are antidemocratic (Brown 2004, Pirtle 2007: 66–73).

The following will discuss the NASA Asteroid Initiative pTA as an example of how social science techniques can both be valuable and a joint effort between philosophers and social scientists. We will sketch out how the results from this forum raise questions that philosophers and reflective practitioners should engage in order to improve pTA and make it more useful in future applications. Indeed, the assessment approach here also questions to what extent Kitcher is right in dismissing how WOS could be implemented in small but important localized situations.

Participatory Technology Assessment

What follows is a case study of a pTA used to inform engineering decision making at a federal agency, specifically NASA's Asteroid Redirect Mission, which is a subset of the broader Asteroid Initiative. This pTA, which needs to be viewed in historical context. Technology assessment can be generally seen as a set of systematic analysis approaches for assessing how new technologies and engineering systems will affect and be affected by society (Sclove 2010). As Matthew Wisnioski describes in his book, *Engineering for Change* (2012), the origin of technology assessment can be traced to concerns about technology and engineering during the 1960s. As the Vietnam and Cold Wars brought forth ever more advanced military systems and a lingering threat to society due to nuclear weapons, questions about whether technology was generally a force for good became more common. Many engineering societies formed 'Technology and Society' committees as sites for dialog about the proper role for engineers in society. Increasing awareness of the unintended consequences of technology led Congress to establish the U.S. Office of Technology Assessment (OTA) in 1972. Until its disbandment in 1994 due to budget cuts, OTA performed a series of technical analyses for Congress using a range of assessment techniques. While OTA has been criticized for various reasons, including not engaging the academic community doing research on how to assess technology, it did help formalize tools and create interest that fueled a larger technology assessment community (Morgan and Peha 2003). Technology assessment work has helped to build on an old and continuing set of tools for assessing the societal value of various research and development efforts (Guston 2000; Toulmin 1961; Mowery and Rosenberg 1979; Coates et al. 2001; Bozeman and Sarewitz 2011). The need for such engagement is still clear: despite all of this work, there is no systematic and agreed-upon method in the government for thinking through the noneconomic values of research (Bozeman and Sarewitz 2011).

Participatory technology assessment is distinct from other types of technology assessment, such as parliamentary technology assessment, constructive technology assessment, and real-time technology assessment (Schot and Rip 1997; Guston and Sarewitz 2002). Participatory technology assessment differs due to its focus on engaging the public in making systematic assessments of technology, seeking to empower the public to consider decisions that some might otherwise think a lay public would be incapable of doing. The broad focus of pTA is on assessing values and helping to inform decisions. It can be difficult to get conclusive results from pTA that dictate what policy results should be, but, by going through the pTA process and obtaining results, managers are able to reflect better on their own values, and may be influenced to make decisions that better achieve what the public wants.

As will be seen in the description of the methodology that follows, pTA in some ways bears a strong resemblance to the WOS thought experiment described by Kitcher. It brings forward as practically representative a group of citizens as possible, allows them to learn about a topic, and provides these citizens an opportunity to give recommendations to decision makers (Tomblin et al. 2015b). However, the pTA under consideration focused on specific issues and decisions, not on a full consideration of all of science and engineering priorities. It also did not distinguish among the three facets of WOS: funding priorities, proper investigation, and desired application. The goals of the Asteroid Redirect Mission and the methods of how to achieve it were discussed at the same time. In part, the pTA can be seen as an experiment in collapsing the levels of WOS, focusing on actual decisions.

pTA of NASA's Asteroid Redirect Mission

NASA partnered with an outside group, the Expert and Citizen Assessment of Science and Technology (ECAST) network, to examine NASA's Asteroid Initiative. Formally titled "Informing NASA's Asteroid Initiative: A Citizens' Forum," two pTA forums were held in November 2014, one in Phoenix, Arizona and one in Boston, Massachusetts. Deliberation themes included the full Asteroid Initiative, including asteroid detection, planetary defense, and the planned Asteroid Redirect Mission (ARM), as well as the broader Journey to Mars for human exploration. Here, we recount results from the ARM deliberation only, which was the portion of the larger deliberation most closely tied to decision making.

Informing ARM with public input took place at an important time, as NASA continues to plan its human space flight agenda. The ARM project is an ongoing development with an aim to redirect an asteroid into lunar orbit where astronauts would rendezvous with it. The mission is a major agency priority, helping to accomplish President Barack Obama's goal for human exploration of asteroids. In early planning for ARM in 2014, NASA announced that it would choose between two technological options, to 'downselect' on a final mission concept, as discussed later in this chapter. For ARM, as with all of its missions, NASA wants its engineering work to benefit society, and sought out this public deliberation so as to be better informed when making decisions. The goal was for NASA to get structured outside input prior to making decisions, such that the information might better enable NASA to implement a solution that is best in keeping with what society wants. The ARM deliberation served as an experiment in using qualitative public input to inform engineering decision making, as the results were shared with NASA managers prior to making key technical decisions.

This instantiation of pTA brought a group of citizens together over the course of a day in the two locations. The ECAST partners organized and ran the forums, including the selection process for who could attend. Ninety-seven citizens attended the forum in Phoenix and eighty-seven in Boston. At each forum, attendees were organized into deliberative groups of six to eight people who focused on common technical and social issues together. Attendees represented broad demographic diversity (described in Tomblin et al. 2015b), covering a range of ages, economic backgrounds, ethnicities, and educational backgrounds. ECAST deliberately selected forum participants to make the participant pool as 'neutral' as possible, minimizing the proportion of traditional NASA stakeholders to a proportion found in the general U.S. population. Further, NASA personnel were not allowed to verbally interact with participants so as to ensure that they would not influence the discussion.

ECAST carefully structured the level of public 'informed-ness' about the Asteroid Initiative. A week prior to and at the meeting, ECAST provided participants with written background material about the initiative. During the event, this background information was reinforced with a video narration of the material. Background content was developed through dialog between

NASA and ECAST, with ECAST taking responsibility for ensuring it represented a balanced view of all issues and provided sufficient information to make informed decisions. Facilitators lead each group in deliberations, maintaining fair and balanced dialog among participants and providing clarifications on basic technical questions concerning the background material. For technical questions that exceeded the grasp of the facilitators, citizens sent an electronic question to NASA experts, who quickly responded. During the event, public input was collected in three ways: (1) participant written statements and voting; (2) researchers observing and audio recording the social dynamics of four groups at each site; (3) and pre- and postsurveys that measured participant motivation, prior knowledge, attitude change, and perceptions of the forums.

Deliberation Topic: The Asteroid Redirect Mission Downselect Between Option A vs. Option B

Using the deliberation approach described, the technical decision citizens discussed was about two different options for accomplishing ARM. In short, the mission as originally proposed is what eventually became called Option A. The approach involved sending a vehicle, the Asteroid Robotic Redirection Vehicle (ARRV), to dock with a small (10 meters in diameter) asteroid, to envelop the asteroid in a bag, and then use a constant propulsive force to control the trajectory of the asteroid and slowly put it in the moon's orbit over the course of several years. Once there, the crewed portion of ARM will occur. Astronauts on an Orion spacecraft will rendezvous with the ARRV and the captured asteroid. Once docked, the astronauts would unfold the bag, collect samples, and conduct other research on the asteroid. NASA's plans for this mission continued to evolve after it was first announced (Muirhead 2014; Mahoney et al. 2014; Gates et al. 2014). The desire for greater certainty on the structure of the target asteroid, as well as desire for more candidate options, led to Option B being developed (Mazanek 2014). This option involved traveling to a larger asteroid, greater than 100 meters in diameter. Previous probes have shown that large asteroids have numerous boulders resting on their surface that are loosely held in place by a slight gravitational attraction. The ARM Option B probe would descend on the asteroid, landing just above a boulder. Robotic arms would grab the boulder, secure it, and the probe would then move the boulder to a distant retrograde orbit around the moon, where astronauts could dock with it.

The pTA participants were asked to choose between the two options, and were provided some of the same technical information given to NASA managers. The following summarizes the background information given to participants at the forum:

- Technology development: At a top level, both options would develop solar electric propulsion (SEP), which was a primary goal for the mission. SEP would make possible sending large volumes of payload to Mars with significantly less propellant. Options A and B were equally beneficial in this respect.
- Sample size and composition: Option A would ostensibly retrieve the larger sample, but challenges exist in being able to accurately gauge the physical composition and size of an asteroid prior to choosing a sample to capture. The mission potentially could retrieve a sample less than half the size of the targeted 10-meter diameter. In comparison, Option B would involve challenges with removing the boulder from the surface of the asteroid, but would have higher confidence on the composition and size.
- Planetary defense: Both options would provide benefits for helping with planetary defense. Option A could utilize an ion beam deflector and a so-called gravity tractor maneuver. Option B could use the additional mass of the boulder to do what is known as an 'enhanced

gravity tractor,' where its influence would be significantly enhanced by the mass of the boulder. Both approaches could be used in future missions to deflect asteroid threats.

- Mars extensibility: Option B was envisioned as having a use extensible to going to Mars, as it could be used to retrieve samples from the moons of Mars (Phobos and Deimos).
- Orbital debris: Option A has been discussed as having potential dual uses for orbital debris removal from Earth orbit.
- Cost: Initial NASA estimates showed that the cost of both options would essentially be the same, with both costing less than \$1.25 billion, meaning that cost was not a deciding factor between the two options.

Overall, the ARM downselect had uncertain dimensions to the decision that could be affected by what values and preferences the public and decision makers hold.

Results on ARM Option A vs. Option B

In this section, selected results on the ARM deliberation from the ECAS report on the Asteroid Initiative are discussed (Tomblin et al. 2015a; Tomblin et al. in press). One qualitative result speaks well about the quality of the data collected overall: while some participants initially struggled with the technical aspects of the topic, overall groups had strong and engaging discussions, with observers noting that participants were able to process and comment on substantive technical issues.

The first set of quantitative results includes the voting preferences for the ARM mission itself. Individuals voted both as a group (which required a consensus decision) and individually. Option B was the overall winner, with 78% of the total table votes going toward it. Option A had four groups out of twenty-seven vote for it. Two tables in Phoenix abstained from making a decision, partly because they were unsure of whether the private sector should have a larger role in this work. Individual votes were roughly similar to the group vote totals.

Beyond looking at the vote totals, participant rationales detailing their reasons for supporting a given choice were also collected. We coded participant responses and lumped them into major categories that underlay their decisions in favor of either Option A or Option B. These results, which follow, can be read as the most common rationales or values that guided the technical decision to go with ARM Option A or Option B. These themes in some cases represent values, such as valuing extensibility to go to Mars, and in other cases represent perceived risks, such as the perception that the option would likely be successful. In parentheses we report the percentage of times that the rationale appeared in participant statements related to a particular theme for either Option A or B. Most rationales were highly related to Option B, with collecting space junk the only rationale favoring Option A:

- Science: the way either option would advance usable knowledge about space (Option B = 87.3%)
- Technology: emphasis placed on technological advance to explore deep space and its resources (Option B = 87.1%)
- Asteroid sample: an interest in getting a larger sample of an asteroid, implicitly to get greater economic or scientific value (Option B = 64.9%)
- Potential: the promise that either option had in leading to even more scientific and technological advances (Option B = 82.4%)
- Success: the degree to which participants believed either option would succeed (Option B = 75.8%)
- Control: the perception that controlling objects is key to learning something from asteroid samples (Option B = 77.4%)

- Going to Mars: the option was seen as desirable for the benefits it would have in furthering technologies needed to explore Mars (Option B = 96.6%)
- Gravity tractor: desire for the enhanced asteroid deflection approach that could be demonstrated by Option B, in furtherance of planetary defense (Option B = 100%)
- Exploration: belief in whether an option would lead to more exploration (Option B = 96.0%)
- Planetary defense: a generic interest in furthering planetary defense goals (Option B = 100%)
- Flexible: a notion that the option can serve multiple purposes (Option B = 90.9%)
- Mining: desire to economically take advantage of the final sample (Option B = 77.8%)
- Collecting Space Junk/Debris: interest in a secondary ability of Option A to help capture and dispose of excess debris in space (Option A = 91.7%)

Generally, those who preferred Option A felt it would collect a larger asteroid sample that could have greater scientific and economic benefits because of its size. Common themes for Option B include Mars extensibility, greater certainty in the type of boulder (or asteroid material) that would be obtained, and a greater perception that it will benefit planetary defense. When presented to NASA management, the focus of the Option A vs. B votes was less on the vote count but on the common rationales provided for the decisions. Several managers thought that the results were “not surprising” and found value in the results. The iterative dialog between ECAST and NASA did generate a rough public value map that highlighted the socio-technical complexity that lay citizens introduce into an assessment of emerging technology that NASA managers could reflect on in relation to their decision about Option A vs. B. Once questions about the legitimacy of the “Informing NASA’s Asteroid Initiative” forums were addressed, it did allow internal NASA briefings to be focused more on the diverse and at times conflicting values expressed in different participant rationales.

Results on What ARM’s Goals Should Be

Participants were asked to provide what they thought NASA’s goals should be as it goes about planning the ARM mission, regardless of whether it pursues Option A or Option B. The results are shown in Table 29.2, which shows the ranked goals of the asteroid mission, essentially providing input on the goals that could be used to make a downselect decision.

The top three goals—science, planetary defense, and technology development for human space flight—were essentially tied as the top result. These are significantly higher priorities

Table 29.2 Participant average priority rankings of seven potential goals* of the Asteroid Redirect Mission on a scale of 1–7 (1 = highest priority; 7 = lowest priority). Note: Standard deviations are reported in parentheses.

<i>Goal</i>	<i>Boston</i>	<i>Phoenix</i>	<i>Combined</i>
<i>Advancing science</i>	2.76 (1.40)	2.51 (1.78)	2.63 (1.62)
<i>Advancing planetary defense</i>	2.54 (1.81)	2.85 (1.63)	2.71 (1.72)
<i>Advancing technology needed for human spaceflight</i>	2.65 (1.40)	2.87 (1.78)	2.77 (1.62)
<i>Redirecting an asteroid that no one has been to before</i>	4.01 (1.82)	4.71 (1.73)	4.38 (1.81)
<i>Developing the economic potential of asteroids</i>	4.67 (1.45)	4.36 (1.85)	4.51 (1.69)
<i>Engaging with commercial and international partners</i>	5.00 (1.54)	4.86 (1.56)	4.93 (1.55)
<i>Performing an exciting mission</i>	6.05 (1.26)	5.68 (1.71)	5.86 (1.52)

* The goals are arranged in ascending order from highest to lowest priority based on voting in Massachusetts and Arizona combined. The voting in Massachusetts and Arizona are fairly consistent. Table and description reproduced from Tomblin et al. 2015b with permission.

than the other goals that were listed, including economic potential. It has long been known that planetary defense is an oft-ignored priority (NRC 2010). One key piece of education occurred in the forum during the morning sessions on asteroid detection and planetary defense, which provided general statistics about asteroid threats, possible impact scenarios, and mitigation options for humans to deal with them. Given all of this information, the preference of the public to rank planetary defense as equal (not greater, or less) to science and technology development of human space flight may be seen as informed. This suggests that planetary defense should be an important topic of conversation about the goals for future NASA mission planning.

The implications of this analysis of goals could be varied. Specifically, for ARM, if a decision maker—complete with the best technical understanding of the risks involved in the downselect—was able to see greater benefits for one of the ARM options in supporting these top goals, then these results could influence or change how they decide on the final decision. For example, if the two ARM options had equal benefit for technology development for human spaceflight and science, but one had greater benefit for planetary defense, then that might potentially sway or influence the decision.

All of these results were presented to NASA managers prior to NASA making its 2015 decision between Option A and Option B. While all aspects of NASA's final decision in favor of Option B are not public, we did ask NASA managers how useful the results were and how they played a role in the decision. Accordingly, the input on goals for ARM was said by managers to be some of the most useful results. Further research is necessary to assess what the right way to engage the public on goals for research should be. The level of granularity on goals solicited here was at a high level, meaning of cotozems gave input on lower level goals (such as different science objectives, economic benefit to different actors) it would require new interpretation.

Conclusion

Philosophers have worked out many ways to think about the relationship between science and democracy (Longino 1990; Douglas 2009). Kitcher's WOS is a prominent example and aligns in interesting ways with the NASA pTA experiment described here. As discussed in this chapter, there are challenges in Kitcher's framing, including his presumption that philosophy should assemble an ideal that is then to be built upon by future social science work. The pTA example here and the history of technology assessment generally shows that there have long been social science tools available for thinking through the role of science in democracy. Moreover, there are interesting parallels between the pTA method and the steps of Kitcher's WOS, showing that it may be possible to implement in localized cases using representative and informed groups. In contrast to building off of an ideal, it is possible to 'jump in' and implement a social science technique, pTA, that actively informs interesting conceptual issues during policy formation with 'neutral' public input.

The pTA analysis examined a vital early part of NASA's human exploration plans, the Asteroid Redirect Mission, and was used to inform a major decision. How do this and potential follow-on activities align with Kitcher's WOS? WOS would require thinking through and contrasting how much resources go to different areas of science and engineering, such as how much society should support health vs. space research. In this deliberation topic, we focused only on assessing questions within an existing funding stream, on human space flight and ARM, and did not ask participants to actively compare competing priorities. However, we could have done so—and the current pTA results still offer potential insights on how competing priorities may need to be addressed at NASA. If one takes seriously the participants' emphasis of planetary defense as being a top goal for ARM, perhaps there are other aspects of NASA's planetary defense work that could be better developed to address public needs. Additional pTA activities

could be implemented to compare how much funding should go to different agencies. If the rigor of the analysis is greater, then prospective pTA activities could potentially involve citizens for much longer periods of time. Some Danish Consensus Conference involve citizens over the course of months to assess a given topic, which would be ideal for a multiagency comparison (Sclove 2010).

However, despite the need to scale up the participatory engagement to cross-disciplinary and cross-agency purposes to develop a closer analog to WOS, there is a very real sense in which a pTA-driven exploration infrastructure could achieve the core intent of Kitcher's argument. Support for the implementation of pTA could be supported by many of his philosophical premises. Implementing pTA as a social science method and not as an ideal would change what the focus of philosophical work is, following an approach closer to Brown's view.

Shifting away from ideals need not diminish the role of philosophers or reflective practitioners. By forcing philosophers to engage with actual democratic deliberation that informs policy, many more philosophical nuances emerge that need sorting out, including:

- How do values and technical judgments shape decision rationales? The ARM pTA rationales on the respective values of the ARM Options A and B show a range of nuances about how technical risks and social values affected the votes. Teasing those apart and making them accessible to engineers, managers, and elected officials is a task that requires many different intellectual viewpoints to understand why the public—and eventually NASA managers—make a decision.
- How do managers, engineers, and members of the public differ in terms of their values and thought processes? One could have performed a pTA deliberation comprised of NASA personnel as participants and may likely have encountered diverse views in the ensuing deliberation. What would it mean for the nature of expertise if a NASA-internal pTA did or did not align with the dynamics of this public pTA?
- What informs cultural differences about values? Why did some prefer Option A over Option B? We explore some of that work in Tomblin et al. 2015b, but education levels and attitudes about independence and the private sector seem important.
- How many members of the public should be engaged? Two forums were used for this pTA, with many of the same value statements and thoughts being described at each site. pTA does not try to poll thousands of people to get a statistical generalization, instead provides more informed, deliberated, input (Sclove 2010). It may be the case that a broad sampling of the public's views can be obtained just through two locations of roughly ninety people, but one could have engaged any number of citizen groups to see if important differences emerge across sites. Future work should examine how representative and numerous of a deliberation group one should pursue, hopefully highlighting a practicable minimum threshold for engaging more members of the public.
- How exactly should an agency use the results? In distributing the results through NASA, we focused on presenting the richness of the data, so as to enable NASA decision makers to compare the participants' values and rationales to their own. In some cases, the values expressed by the participants align with NASA's values, and helped to reaffirm NASA's decision criteria (Tomblin et al., in press). But for other pieces of data, managers said they benefited from considering the results but could not clearly say how they would use the results other than as a tool for reflection.
- When should the public decide? Lurking behind all of this lies a basic truth: despite the engagement here, NASA managers still made the final ARM downselect decision. At what point, if any, should the results of a pTA exercise actually be used to dictate policy? One could experiment with letting pTA exercises make certain technical decisions, and

empirically see if the results have value over the long run. At minimum, pTA can help decision makers at agencies and in Congress to be more aware of their own thoughts as they make decisions.

- How do we separate means and ends? It is not clear if the deliberation about ultimate goals vs. the assignment of methods on how to accomplish goals can be neatly demarcated, as Kitcher's WOS assumes. Our pTA discussed two means to accomplish ARM, and assessed ends and means simultaneously. Given that policy processes often have constrained decision-spaces, such a mixing of ends and means may be necessary.

Overall, the NASA pTA effort marks the first time a U.S. federal agency has engaged in participatory technology assessment, and offers an interesting example for philosophers to reflect on. The use of a pTA forum to inform a very specific systems engineering and architecture decision is likely also unprecedented. This pTA effort reinforces other findings in the literature that show the public is able to have rich discussions about technical topics and to produce valuable insights for decision makers (Stirling 2008).

While the ARM decision has been made, future decisions and deliberations await.² The success of the public in deliberating on ARM indicates that the public might be able to meaningfully iterate on major decisions for future space architectures. This raises the possibility that a series of major decisions could be discussed and assessed in this way, increasingly leading to something closer to Kitcher's vision of well-ordered science. What would a pTA-driven exploration architecture look like, where public values, deliberation and perspectives are infused at each step of the process? (See Kaminski 2012 for additional thoughts here.) Would the architecture that follows from such a process really be better aligned with what society wants and needs? How would such an architecture fare at dealing with the myriad of space policy stakeholders, interests, and time frames that are needed to accomplish an ambitious mission? More data and experimentation with social science methods like pTA are needed before answering these questions. Philosophers can play a role in unpacking and evaluating these questions, but only if they dig in and help to tease apart the complex sociotechnical nuances that can be only be revealed by having diverse conversations such as those created by pTA.

Note

- 1 All opinions in this paper are our own and do not necessarily reflect the views of the National Aeronautics and Space Administration or the U.S. Government.
- 2 Since this chapter was written, the NASA President's Budget Request proposed to cancel the Asteroid Redirect Mission and the program is currently closing out its activities. Much of the work of ARM will continue to be used in future NASA programs.

References

- Bozeman, B., and Sarewitz, D. (2011). Public Value Mapping and Science Policy Evaluation. *Minerva*, 49(1), 1–23.
- Brown, M. B. (2004). The Political Philosophy of Science Policy. *Minerva*, 42(1), 77–95.
- Coates, V., Farooque, M., Klavans, R., Lapid, K., Linstone, H. A., Pistorius, C., and Porter, A. L. (2001). On the Future of Technological Forecasting. *Technological Forecasting and Social Change*, 67(1), 1–17.
- Douglas, H., 2009. *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh.
- Gates, M., and Muirhead, B., Naasz, B., McDonald, M., Mazanek, D., Stich, S., Chodas, P., and Reuter, J. (2014). NASA's Asteroid Redirect Mission Concept Development Summary. Paper given at *IEEE Aerospace*, March 8, Big Sky, Montana. 978-1-4799-5380-6/15.
- Guston, D. H. (2000). *Between Politics and Science: Assuring the Integrity and Productivity of Research*. Cambridge University Press.

- Guston, D. H., and Sarewitz, D. (2002). Real-Time Technology Assessment. *Technology in Society*, 24(1), 93–109.
- Kaminski, A. P. (2012). Can the Demos Make a Difference? Prospects for Participatory Democracy in Shaping the Future Course of U.S. Space Exploration. *Space Policy*, 28(4), 225–233.
- Kitcher, P. 2001. *Science, Truth, and Democracy*. Oxford: Oxford University Press.
- Kitcher, P. 2011. *Science in a Democratic Society*. New York: Prometheus Books.
- Longino, H. E. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.
- Mahoney, E., and Herrmann, N., and Gates, N. (2014). An Overview of NASA's Asteroid Exploration Efforts: Past and Present. Paper presented at the *65th International Astronautical Congress*, Toronto, Canada. Available at: www.nasa.gov/sites/default/files/files/IAC-14B31x24563-Mahoney.pdf
- Mazanek, D. (2014). Asteroid Redirect Mission (ARM) Robotic Boulder Capture Option (Option B). Presentation given at the *Small Bodies Assessment Group*, July 30 meeting in Washington, DC. Available at: www.lpi.usra.edu/sbag/meetings/jul2014/presentations/1030_Wed_Mazanek_ARM_OptionB.pdf
- Morgan, G., and Peha, J. (eds.) (2003). *Science and Technology Advice for Congress*. Washington, DC: Resources for the Future Press.
- Mowery, D., and Rosenberg, N. (1979). The Influence of Market Demand Upon Innovation: A Critical Review of Some Recent Empirical Studies. *Research Policy*, 8(2), 102–153.
- Muirhead, B. (2014). Asteroid Redirect Robotic Mission (ARRM) Concept Overview: Briefing to SBAG. Presentation given at the *Small Bodies Assessment Group*, July 30 meeting in Washington, DC. Available at: www.lpi.usra.edu/sbag/meetings/jul2014/presentations/1030_Wed_Muirhead_ARM_OptionA.pdf
- National Research Council (NRC), Committee to Review Near-Earth Object Surveys and Hazard Mitigation Strategies. (2010). *Defending Planet Earth: Near-Earth Object Surveys and Hazard Mitigation Strategies: Final Report*. Washington, DC: National Academies Press.
- Pirtle, Z. (2007). *Democratizing Nanotechnology: Intersecting the Philosophy of Science and Science Policy*. Honors thesis at Arizona State University.
- Schot, J., and Rip, A. (1997). The Past and Future of Constructive Technology Assessment. *Technological Forecasting and Social Change*, 54(2), 251–268.
- Sclove, R. (2010). *Reinventing Technology Assessment: A 21st Century Model*. Published by the Science Technology Innovation Program. Available at: www.loka.org/documents/reinventingtechnologyassessment1.pdf
- Solomon, M. (1994). Social Empiricism. *Nous*, 28(3), 325–343.
- Stirling, A. (2008). “Opening Up” and “Closing Down” Power, Participation, and Pluralism in the Social Appraisal of Technology. *Science, Technology & Human Values*, 33(2), 262–294.
- Tomblin, D. Pirtle, Z., Farouque, M. Sittenfeld, D., Mahoney, E., Worthington, R., Gano, G., Gates, M., Bennett, I., Kessler, J., Kaminski, A., Lloyd, J., and Guston, D. In press. “Integrating Public Deliberation into Engineering Systems: Participatory Technology Assessment of NASA's Asteroid Redirect Mission” *Astropolitics*.
- Tomblin, David, Worthington, R., Gano, G. L., Farouque, M., Sittenfeld, D. and Lloyd, J. (2015b). *Informing NASA's Asteroid Initiative—A Citizens' Forum. Final Results Report*. Published by Expert and Citizen Assessment of Science and Technology (ECAST) network. Available at: <https://ecastnetwork.files.wordpress.com/2015/09/nasa-asteroid-initiative-citizen-forum-full-report.pdf>
- Tomblin, D., Worthington, R., Gano, G., Farouque, M., Sittenfeld, D., and Lloyd, J. (2015a). *Informing NASA's Asteroid Initiative—A Citizens' Forum*. Published by Expert and Citizen Assessment of Science and Technology (ECAST) network. Available at: www.nasa.gov/sites/default/files/atoms/files/ecast-informing-nasa-asteroid-initiative_tagged.pdf
- Toulmin, S. E. (1961). Foresight and Understanding: An Enquiry Into the Aims of Science.
- Wisnioski, M. (2012). *Engineers for Change: Competing Visions of Technology in 1960s America*. Cambridge, MA: MIT Press.

When Loving Your Planet Isn't Enough

*Diana P. Hoyt*¹

There can be no doubt that science and technology are inextricably part of society and human social enterprise. From the beginnings of science to the industrial revolution, scientific discovery dramatically extended human understanding of the world, and technology our grasp. In the late twentieth century, work in the fields of philosophy of science and technology demonstrate that just as science and technology shape society, the social enterprise also shapes science and technology, from the design and certification of scientific evidence, to the creation and evolution of artifacts and technology. Scholars demonstrate that human values saturate science and technology, an influence made visible by scholarship in fields as diverse as philosophy, sociology, and history of science and technology. The goal of this chapter is to identify and to briefly characterize a range of issues pertaining to space exploration, each providing the scholar with opportunity to explore social and epistemic dimensions of the exploration enterprise.

The use of the 'frontier' metaphor in the U.S. space program creates a reality (and ontology) that hides important ethical issues embedded in the shape and configuration of the U.S. space program itself as well as its underlying values. Exploring ethical entailments of these metaphorical inducements raises significant public policy issues in space exploration. These include, but are not limited to: Shall space resources be consumed, by whom, and how? Do Martian microbes have moral standing? Is terraforming a planet an ethical act?

Currently, these issues remain hidden from public discussion by virtue of the long-standing incorporation of the metaphor of the frontier in the U.S. space program. Specifically the metaphor of the frontier creates our understanding of outer space, shapes the goals of the U.S. space program, and is reflected historically and currently in the configuration of NASA's program of space exploration.

For example, the use of the frontier metaphor in space exploration emphasizes the concept that space explorers are somehow larger than life, heroic in character and deed, and cut from the same cloth as American explorers and pioneers of the eighteenth and nineteenth centuries. NASA and the space program draw their inspiration for the notion of the frontier largely from the works of Frederick Jackson Turner, who on July 12, 1893 presented his paper entitled "The Frontier in American History" to a meeting of the American Historical Association in Chicago. The paper, which focused on the influence of the frontier in the shaping of the United States and the American character, influenced generations of historians. Profoundly embedded in Turner's description and analysis was the notion of American exceptionalism, which was enthusiastically embraced by the space program, in rhetoric and in program configuration.

The use of Turner's exceptionalism by NASA promotes a 'frontiersman-like' approach to the natural resources of the cosmos—it encourages the notion that extraterrestrial resources are humanity's for the taking. In opposition to the conservation ethic embedded in the environmental movement, this public policy position is neither noticed, discussed, nor negotiated, for it is a natural consequence of the way in which the notion of the frontier of space organizes our

concepts of space exploration in terms of utilization and consumption, instead of preservation and conservation.

Using the frontier metaphor to explain or justify space exploration entails the adoption of the following stance: if terraforming advances exploration, then terraforming (the radical alteration of planetary climates to create different atmospheres) is a logical and acceptable exploration strategy. Under this scenario, neither the planet in question nor the flora, fauna, or strange creatures, if any, that inhabit it, have standing or rights. This is an utterly species-centric argument and one that, in the final analysis, may indeed carry the day. But is also one that, in an ethical society, ought to be discussed and debated. Not everyone would agree that humanity's ends rule supreme. It may be that this is a perfectly acceptable consequence of exploration, but it is one that merits debate and discussion by society at large. When the metaphor of the frontier is used in conjunction with space exploration, this discussion does not take place, for it is already part of the assumptions inherent in our understanding of exploration occasioned by the use of the frontier metaphor.

The metaphor of the frontier thus has the ability, covertly and surreptitiously, to organize our views of reality, including humanity's understanding of the moral status of 'things' in the cosmos, and it results in public policy consequences of embracing the frontier metaphor without question. The fact that the frontier metaphor applied to space exploration has penetrated our consciousness so deeply in the United States undermines our ability to engage in a discussion of space ethics from Rawls's point of view of the veil of ignorance. The examination of space discourse and its hidden value propositions provide a rich field of enquiry in parsing issues pertaining to space exploration.

Since the launch of the first Earth-orbiting satellite, Sputnik, on October 4, 1957, space exploration has been largely characterized in terms of the enabling technology, such as the development of rockets and propulsion, sensors and satellites, lunar landers, rovers, and telescopes, rather than as a human endeavor in which values play a significant role. With the expansion of the space program to include other planetary bodies, the drive to launch humans well beyond Earth's orbit, to Mars and beyond, has grown rapidly. Further, what was once exclusively the domain of government has been diversified. As a matter of public policy, the U. S. government has turned low Earth orbit over to the private sector to develop and commercialize at will. Companies have been formed to build launch vehicles and space capsules to send humans and cargo to the International Space Station and Mars, mine asteroids and other planetary bodies, and provide opportunities for adventure tourism in outer space. Humans are increasingly turning to outer space for the practice of formerly Earth-based activities such as commerce, resource extraction, mining, and tourism. The early adopters for space commercialization have invested a significant amount of their own money in these enterprises and they come from the ranks of the very wealthy.

Sir Richard Branson (Virgin Galactic) is founder of the Virgin Group; Elon Musk's (Space X) fortune was made by PayPal; Jeff Bezos (Blue Origin) is the founder of Amazon; and Paul Allen's (Vulcan Aerospace) fortune comes from Microsoft. Billionaire entrepreneur Elon Musk disclosed plans to send an unmanned capsule to Mars, perhaps as soon as 2018, revealing the first details of his long-term vision for a private enterprise to colonize the red planet.

When Musk announced that Space X was ready to go to Mars, the reaction in the space community ranged from disbelief to advocacy. An article in NASAWATCH, a nonofficial website devoted to all the news that NASA doesn't publicize, contrasted Musk's commercial space initiative with the NASA program of Mars exploration:

What you have seen this week [April 29, 2016] is a paradigm shift hiding in plain sight. In September Elon Musk is going to reveal his plans for colonizing Mars. This announcement

was just the opening note. A private sector company has committed to spend its own blood and treasure on a mission to another planet. They have not asked NASA for a penny for this mission and have offered to tell NASA what they have learned—for free.

(Cowing 2016)

Musk believes that we must colonize other worlds if the human race is to survive:

I think there is a strong humanitarian argument for making life multi-planetary in order to safeguard the existence of humanity in the event that something catastrophic were to happen, in which case being poor or having a disease would be irrelevant, because humanity would be extinct. It would be like, “Good news, the problems of poverty and disease have been solved, but the bad news is there aren’t any humans left.”

(Anderson 2016)

Musk’s company, Space X, has already successfully delivered cargo to the International Space Station on multiple missions. Bigelow Aerospace was founded by Robert Bigelow and is funded in large part by the profit that Bigelow gained through his ownership of the hotel chain Budget Suites of America. Commercial companies seek to provide a range of capabilities: crew and cargo transport vehicles (27); commercial space stations (6); launch vehicles (24) landers, rovers and probes (20); research craft and technology demonstrators (14); and space mining and manufacturing (10). While many of these are small start-up companies founded by space enthusiasts, the early adopters and major players, however, come from big money. Commercial space activities are still the playground of the elite—in this case, the very rich. The question remains whether or not this trend will evolve over time to include ‘ordinary people’ or whether commercial space activities will be the domain of the privileged few to the exclusion of the majority of humans on planet Earth. This uneven distribution in access to space raises the question of equity.

The expanding presence of humans in space and the continued expansion of human activities into space raise a suite of issues about how humans ought to be in space. These issues are rooted in ethics and philosophy, epistemology, teleology, ontology, anthropology, and sociology, and call for new public debate and discussion on the behaviors of the human species in the space environment. At this writing, there is no agreement on how humans ought to behave in space. The biggest challenge to space exploration as human endeavor will be the creation of a common moral community, able to negotiate common ways of being in space. Questions pertaining to ethics, justice, government, and human obligations to environments, as well as to both biological and nonbiological entities, rights, and religion and theology, emerge. These issues contain and are governed by human thought and values and are unique to the human species. They hinge on understanding of what it means to be human. A great deal of work has been done on the roles of values in society, with a particular focus on social systems and the expression of value-based social solidarities. Scholars conclude that perceptions of science, technology, and environment are part of people’s general orientation towards human relationships and technological worlds, including outer space. If humans are indeed to become an interstellar species, questions concerning values, ethics, and moral action must be discussed openly, intentionally, and publicly if space exploration and exploitation are to be ethically informed enterprises.

All of these topics are linked by questions related to a common frame, and all involve principles of fairness, equity, justice, and power in social relationships. While technology plays a starring role in space exploration, myriad fundamental issues revolve around power and control: how are decisions made, and who makes these decisions? While many of the issues pertaining to space exploration appear straightforward on the surface (should humans litter in space?),

these fundamental questions interrogate the source of power in societies with unequal social distributions of power. In the case of space exploration, for example, the determination of an ethical action framework for exploration will be left to those who are able to explore. Among myriad competing interests, who will determine whose opinion will be counted? To borrow from Sandra Harding, Whose space exploration program is this? Whose technology ethic? The discussion of framing and using power and developing a common ethical regime for space exploration often remains hidden behind debates on the eventual outcomes. It is by addressing some of these questions that power relations become visible and then can be addressed now before norms are determined by who gets there first. Many of the emerging issues in space exploration are related to environmental and safety concerns, while other interesting questions spring from the current effort to 'commercialize' space for profit and for recreation.

Justifying Space Exploration

Survival and Existential Safety

Arguments flow from issues pertaining to resources and survival. From the perspective of resources, since the Earth's resources are finite, space resources will amplify our available resources and are necessary for our long-term survival. There are those who believe that the human race is doomed on planet Earth and that we have gone beyond the point where the environmental damage done by humans to the Earth's ecosystem can be reversed. Proponents of this survival rationale point to the future inability of the Earth to support an ever-expanding population, coupled with the depletion of resources needed to support and maintain civilization. Others project a doomsday scenario of an asteroid crashing into Earth, resulting in catastrophic shifts in the climate and environment, and the eventual extinction of the human species, not unlike the extinction of the dinosaurs. Under these conditions Earth would be uninhabitable. Establishing human settlements on other celestial bodies would assure the survival of the human species, and thus is viewed by some as a moral imperative. There are other predictions about how the human species might end, including by pandemic and by war. There are many ways for the Earth to die: by environmental catastrophes such as disruption of the food chain, climate change, nuclear war, nuclear winter, biowarfare, ocean acidification, and other interventions by humans. General Omar Bradley, a leader of accomplishment and distinguished service commented on the proclivity of the human species to make war in an Armistice Day speech given on November 10, 1948: "The world has achieved brilliance without wisdom, power without conscience. Ours is a world of nuclear giants and ethical infants. We know more about war than we know about peace, more about killing than we know about living" (Bradley 1967). How might one contain this aspect of the human character in the conduct of space exploration and settlement and refrain from bringing war and violence into the outer space environment?

There are dangers to the human species far greater than war. According to our current best theories of physics, eventually, our sun will burn out. As it does, it will expand into a giant star that will engulf the Earth, and in all likelihood, incinerate our planet. If Earth survives, it will emerge as a dead planet, incapable of sustaining life. In this scenario, establishing human colonies in space is a sensible, necessary ethical act. Some claim that we are thus morally obligated to pursue space exploration, space settlements, and interstellar migrations. In *Pale Blue Dot* (1994), Carl Sagan frames this discussion in terms of humanity's ability to survive:

Since, in the long run, every planetary civilization will be endangered by impacts from space, every surviving civilization is obliged to become spacefaring—not because of exploratory or romantic zeal, but for the most practical reason imaginable: staying alive. . . .

If our long-term survival is at stake, we have a basic responsibility to our species to venture to other worlds.

(Sagan:1964, p. 371).

Sagan's argument is not without its counterpoint, however. The argument that colonizing space to preserve the human species when the Earth will no longer support life in effect treats the Earth as a 'disposable planet.' By using the argument that if we can't live on this planet, we can always find another, we essentially default on our obligations to Earth's environment and relinquish our responsibilities as planetary stewards. Moving humanity into space may have merit as an act of exploration, but may be questionable without a common ethical framework by which all space explorers operate.

This debate often poses a fundamental ethical question: should we put our resources into colonizing space to avoid extinction of the human species, or should we dedicate our resources to solve the problems on Earth that we have created? Is creating a 'back-up planet' capable of sustaining human life an ethical move? If we destroy our home planet, is it ethical to take over another planetary body for a human survival 'do-over?' Does having a "plan B" for human survival make sense? Whose sense?

Human Purpose and Shared Human Destiny

Another justification for exploring space is anchored in notions of human destiny. Some claim that *wanderlust*—the desire to explore—is written in the human heart. Space advocate Charles Sheldon describes the notion that human destiny is a common reason for space exploration: "There are those who just feel that mankind is destined to step beyond his earthly bonds just as his ancestors once crawled out of the seas . . . Mankind, or his descendants, may be spread to new places and the race will survive even when the Earth itself is no longer habitable" (Sheldon: 1967, p. 74). Space enthusiasts claim that exploration of outer space satisfies an evolutionary drive to explore and live on worlds not our own, and that this expression of the human spirit is our collective, manifest destiny. Exploration, so the claim goes, is firmly etched in our DNA. Space exploration fulfills our human purpose. The question remains whether this reasoning constitutes a moral imperative.

Extraterrestrial Moral Obligations and Rights

Planetary Protection and Ethics

Planetary protection is the term given to the practice of protecting solar system bodies (i.e., planets, moons, comets, and asteroids) from contamination by Earth life and protecting Earth from possible life forms that may be returned from other solar system bodies. According to NASA, planetary protection is essential for several important reasons: to preserve our ability to study other worlds as they exist in their natural states, to avoid contamination that would obscure our ability to find life elsewhere, if it exists, and to ensure that we take prudent precautions to protect Earth's biosphere in case it does. Typically, planetary protection is divided into two major components: forward contamination, which refers to the biological contamination of explored solar system bodies; and backward contamination, which refers to the biological contamination of Earth as a result of returned extraterrestrial samples. There has been much debate in the scientific community with regard to isolation and containment of samples of the Martian soil that we might bring back. At the extreme, scientists contend that a BSL-4 lab is required. This is the maximum level of safety for containment, and is used to contain extremely

dangerous biological samples such as the Ebola virus. The estimated cost to build a facility like this has been quoted at \$30 million. Questions about appropriate protocols, contamination with microbes and organisms, risks of alien microbes brought back to Earth, and catastrophic events make for a difficult negotiation of issues and ethical analysis.

Space Property Rights and Resources

The debate over whether humans can own space resources and property on celestial bodies is an important one if we are to become a spacefaring people. Without the incentives of private ownership, it is unlikely that this 'final frontier' will be appealing to early space settlers. Jim Benson, the CEO of Space Dev, a private company intent on developing space for commerce, puts it this way in an interview for the PBS program *Voyage to the Milky Way* (1999):

I think it is extremely important to create a precedent for private property rights in space. If we make a claim [on a planetary body] we will have some justification or some standing because we took the risk, we paid the money, we flew our spacecraft, and we analyzed the content and the value of that asteroid. We landed on it. It's ours.

The question remains: can John Q. Public stake a claim for 'land' on the moon? In the free market economy, is a space frontier 'resource grab' ethically acceptable? Are free market principles ethical when applied to the allocation of space resources? Are free market principles ethical, period? Given that incentives will attract the private sector to do business in space, and that we must colonize space to assure the survival of the human species, does this mean that space mining and extraterrestrial resource consumption take on the character of a moral imperative? If so, this would conflict with the notion that the natural landscape and environment of planetary bodies has noninstrumental value. What would be a fair process for adjudicating claims and property rights in space? Finally, the question remains: to whom does space belong—those first to arrive and stake a claim?

Harlan Ellison, a noted science fiction author, highlights the danger of free market economics in space, in an interview in *Voyage to the Milky Way*:

Those who are fleetest and most rapacious are going to get there first and are going to do what they want with it. That's the threat of space travel. That those who get there first will not be the ones we want to carry the banner. The ones with great dreams, they will simply be the people who want to strip mine Venus for its pitch blend.

(PBS 1999)

Currently, there are at least two companies that have formed with the specific purpose of asteroid mining. We know of more than 750 S Class asteroids with a diameter of at least 1 kilometer; scientists estimate that each asteroid would contain 1 billion metric tons of iron, or the equivalent of what we mine from the Earth every year. Are humans entitled to strip-mine asteroids? Mining is a polluting activity, and one that is not easy on humans. Would it not be ethical and in the interests of humans on Earth to shift this activity to outer space? Yet, if we remove mining from the Earth, there will be a significant loss of Earth-based jobs in this industry.

One of the determining questions on ownership of space resources relies on our notion of community. Is our relationship with space based on the concept of a *positive community of ownership*? If so, then each of us owns an equal share of space and its contents. The question remains of how we will allocate these space resources. For example, does everyone on the planet get an equal share of all space resources? If there are four planets, does everyone get a

fourth of a planet? Or does each person get their own planet? Or is it more likely that ownership of resources will be distributed? If humans' relationship to space is one of a *negative community of ownership*, then no one has a *prima facie* claim to space property, and all humans share the common starting point of owning no part of space. This is reminiscent of John Rawls's theory of justice as fairness, originating in a common starting point. In the *Theory of Justice*, Rawls describes an agreement situation that is fair among all the parties to the hypothetical social contract. He assumes that if the parties to the social contract are fairly situated and take all relevant information into account, then the principles they would agree to are also fair. The fairness (and in this case, it is procedural fairness) of the original agreement situation transfers to the principles everyone agrees to. Thus, whatever laws or institutions are required by the principles of justice are also fair. With a starting point of a "veil of ignorance," John Rawls proposes a model of a fair choice in which participants' parties would choose mutually acceptable principles of justice. "Since all are similarly situated, and no one is able to design principles to favor his particular condition, the principles of justice are the result of a fair agreement or bargain" (Rawls 1971, p. 11). In this case, the intrinsic fairness of procedural justice provides a framework shaping community values. This observations leads to the question: What model of community should inform our discussion of property rights in space?

The Space Environment

Terraforming

In theory, humans have the ability to rearrange the atmosphere of other celestial bodies to be able to support human life through a process known as *terraforming*. NASA defines terraforming a planet, moon, or other body as the process of deliberately modifying its atmosphere, temperature, surface topography, or ecology to be similar to the environment of Earth to make it habitable by Earth-like life. Through terraforming, a hostile environment could thus be transformed into one able to support human life. Again, fundamental questions arise about our rights of intervention on a planetary scale, about life on other planets, about our obligations to other generations, and on the use of and risks concerning microbes. Writing in the Fall 1999 issue of *Environmental Ethics*, Robert Sparrow, an Australian philosopher, views terraforming as an act of vandalism:

Using an agent-based virtue ethics, I argue that terraforming reveals in us two serious defects of character. First, it demonstrates that we are suffering from a significant aesthetic insensitivity. We would become cosmic vandals. Second, it involves the sin of hubris. We show ourselves to be suffering from an excessive pride which binds us to our own place in the world. In attempting to shape another planet to our ends, we are seeking to become Gods.

(Sparrow 1999, p. 232)

Sparrow disposes with the complex question of whether the Martian environment has standing or intrinsic value. Through the development of an agent-based framework placed in the context of virtue ethics, Sparrow focuses his attention on the behavior of the agent for determining right behavior: "Virtue ethics directs our concern to the character of agents. It asks us to pay attention to the virtues and vices we display through our actions" (Sparrow 1999, p. 230). He argues that the use of agent-based ethics makes possible a further account of the value of complex inorganic systems that holds they have value by virtue of the character traits they expose

in us. Sparrow is a strong advocate for the argument against terraforming, both on Earth and on other planetary bodies. Noting that terraforming turns humans into “world makers,” Sparrow concludes:

The enthusiasm for terraforming looks particularly damning in light of the past technological disasters on Earth. There is little self reflection going on in the debate about terraforming, which is largely a technical debate about feasibility and methods which allows little room for questions about why we would engage in such a project.

(Sparrow 1999, p. 236).

As an alternative to altering planetary environments, we could breed new species of humans to enable them to survive better in space. Would the creation of the equivalent of astronaut farms be ethical? These would be places where science intervenes in human biology to alter our genetic makeup so that humans would be more suited to space exploration. We could, for example, develop a program to breed for radiation-hardened humans who could more easily survive the hazardous radiation of long-term spaceflight. Should we?

Space Debris

Humans have been littering in space for decades, and the amount of debris humans have created in space can be measured. The moon has often been referred to as a celestial garbage can. Things we have left on the moon include five U.S. flags; twelve pairs of boots; ninety-six bags of urine, feces, and vomit; utility towels, used wet wipes; personal hygiene kits; empty packages of space food and other materials, and more than eighty spacecraft, including rovers, modules and crashed orbiters (Garber 2012). The moon currently hosts more than 400,000 pounds of human-produced material.

The U.S. Space Surveillance Network is the leading space object tracking system in the world, and according to a report issued by the Congressional Research Service, the network catalogues objects as small as 10 centimeters in low Earth orbit to and as small as 1 meter in geosynchronous orbit (Hildreth 2014). Estimates of the amount of space junk orbiting planet Earth count 23,000 pieces of debris, ranging from lost tools to abandoned satellites, used rocket stages, and fragments of space hardware. Orbital space debris poses significant threats to space operations and endangers spacecraft of all kinds. Moving at 17,500 miles per hour, a small piece of debris about the size of a marble can destroy a spacecraft or put a hole in the International Space Station.

Intentionally dumping trash in space is not unusual: for decades, space mission profiles called for crash-landings and were a typical method of bringing unmanned lunar missions—and unmanned planetary missions in general—to a close. NASA's typical end-of-mission method has resulted in a trail of debris on planets across our solar system. In addition to pieces of spacecraft, smaller pieces of detritus are also present: these objects were left because they had served their purpose and outlived their utility to their respective missions. Thus, while we regularly leave trash on the moon, much of this trash as well as the debris in orbit around the Earth can be characterized as *strategic trash*, jettisoned intentionally as part of routine mission configuration. Because space launch costs are high, at approximately \$10,000 per pound it was most cost effective to abandon hardware and waste in place. We have not yet developed the equivalent of space garbage trucks to collect debris left in low Earth orbit or on planetary bodies. While debris in orbit eventually reaches Earth as the orbits decay over time, junk left on the moon and Mars bears eternal witness to the human tendency to tread harshly on the world.

Values

Given that it is incumbent on the human species to establish guidelines on how to ‘be’ and act in space, these guidelines must be communally derived from an agreed-upon prioritized system of values and operating principles with regard to what we should and shouldn’t do in space. This challenge poses great difficulty, for values are socially determined, culturally contingent, and can vary significantly among and between cultures and geographies. Steve Rayner of Oxford University concludes that variation in public values can be accounted for more effectively by sociocultural variables than by traditional demographic variables. Values, stabilized by social solidarities (such as membership in organizations like the Sierra Club or the Association of Used Car Dealers), can be understood as frameworks for justification of actions and public policies. When studied, value discourse (for example, the environment as fragile) will yield a commitment linked to a predictable package of preferences for policy goals (Kempton 2001). Given the lack of agreement on values in the environment on Earth, for example, how might one reach consensus on values pertaining to space exploration on the basis of which one develops a uniform extraterrestrial space ethic?

One alternative is to agree on a common approach grounded in a particular way of looking at the world. For example, one might use the lens of pragmatism: the space environment is valuable because we can use it for industry and commerce. Another might be aesthetic: the space environment is valuable because of its stark beauty. A third might be rooted in philosophy: the space environment is valuable because it provides a venue for the fulfillment of human destiny to explore new frontiers.

Worth and Ethical Frameworks

The question remains whether there is a fundamental reason to protect the space environment. Writing for the journal *Space Policy* in 2003, Mark Williamson pointed out that if the space environment has value, from any of these perspectives, then it is worthy of protection. Protection of the space environment demands that humans take action to do no harm to the space environment. Williamson (2004) and other scholars call for the development of space ethics which would cover: “the impact of our action in space on each other, on each other’s property on the Earth, and on the space environment itself.” One of the requirements of a framework for space ethics would involve developing criteria for assessing harm and for balancing interests. While an ethical code has its roots in philosophy and reasoning about moral values, in this case, we need a practicum with tangible consequences.

Margaret McLean, director of bioethics in the Markkula Center for Applied Ethics at Santa Clara University, offers three actionable value-based principles for the conduct of space exploration and commercialization in a February 2006 article “Ethical Considerations for Space Exploration,” posted on the Markkula Center’s website. These are: *space preservation* (space is valued for its own sake, regardless of any benefits that may be derived from it); *space conservation* (we protect the universe’s resources for the sake of others and avoid exploiting these resources for the benefit of a few); and *space stewardship* (we hold ourselves accountable for managing space resources) (McLean, 2016). This last approach mandates that we consider how our actions affect our environment, others, and the future.

Theories for the development of an ethical framework for space exploration adopt a central moral principle from which to define ethical value. These may launch from a basis of intrinsic value or instrumental value. Intrinsic value is defined as the value of an object independent of a valuer. Under the regime of *intrinsic value*, an object has rights and is entitled to justice and respect. *Instrumental value* is that value which requires a valuer, and thus objects with

instrumental value have no rights, and can easily serve as a means to a justifiable end. Martin Fogg develops value theories from four primary 'isms': *anthropocentrism* (only humans have rights, the rest of nature has no moral standing); *zoocentrism* (animals have intrinsic value, and thus moral standing); *ecocentrism* (all life is sacred and has a right to flourish, and ecosystems have intrinsic value, and humans are "just plain biotic citizens"); and *preservationism* (the cosmos has its own values, and thus a right to be preserved without human intervention). Internally coherent moral principles and alternative ethical regimes can be derived for each of these perspectives. Each will yield different practices and consequences.

Risks

Long-distance space exploration could expose astronauts and citizen explorers to high or even unknown health hazards. Many of these risks are poorly characterized, uncertain, unknown, or unknowable. Can astronauts give informed consent to an unknowable suite of risks to human safety and survival? According to a study done by the Institute of Medicine of the National Research Council, long-term exposure in space could place astronauts beyond the limits of 'safe exposure.' For example, the radiation that the Curiosity Rover was exposed to on its way to Mars could potentially exceed the acceptable radiation limits for astronauts if a round-trip journey were made. Under these circumstances, should NASA and other space agencies then reassess the standards and relax the requirements? According to the Institute of Medicine: "The committee finds relaxing (or liberalizing) current health standards to allow for specific long duration and exploration missions to be ethically unacceptable" (National Research Council 2014). Instead, the committee recommended that exceptions to the current health standards be granted on a case-by-case basis and used only in limited circumstances, using the ethics framework developed by the Institutes as guidance.

Planetary and Interstellar Settlements

Colonizing other planetary bodies on the way to establishing interstellar settlements poses problems of 'right behaviors.' For example, the introduction of aliens in the form of humans to other planetary bodies runs counter to our practice on Earth of avoiding the introduction of nonnative species to natural environments. Establishing human colonies in space also raises the question of intergenerational rights and obligations. Children born in space will never experience life on Earth: they will never see the sun rise or experience the feeling of wind on their cheeks. This raises the issue of informed consent: it may not be ethical to deprive the unborn of their rights to experience Earth, which is their native planet. A common argument against space exploration focuses on the fact that as yet, we humans have not solved critical fundamental problems on Earth. There are populations without access to food, clean water, and energy. It would therefore be irresponsible, from this perspective, to spend money on exploring space and colonizing other worlds when we have not lived up to our responsibilities to our fellow humans on this planet.

Justice and Governance

Currently, there are no international laws governing the moon or the protection of the space environment. The Moon Treaty, created in 1979 by the United Nations, declares that the moon shall be developed to benefit all nations, and that no military bases can be placed on the moon. It also prohibits altering the environment of all celestial bodies. To date, no spacefaring nation has ratified this treaty. The Outer Space Treaty, formally the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other

Celestial Bodies, forms the basis of international space law. The treaty was opened for signature in the United States, the United Kingdom, and the Soviet Union on January 27, 1967, and entered into force on October 10, 1967. Among its principles, the Treaty bars states party to the treaty from placing weapons of mass destruction in Earth orbit, installing them on the moon or any other celestial body, or otherwise stationing them in outer space. It exclusively limits the use of the moon and other celestial bodies to peaceful purposes and expressly prohibits their use for testing weapons of any kind, conducting military maneuvers, or establishing military bases, installations, and fortifications. However, the Treaty does not prohibit the placement of conventional weapons in orbit. The treaty also states that the exploration of outer space shall be done to benefit all countries and shall be free for exploration and use by all the states.

The Treaty explicitly forbids any government from claiming a celestial resource such as the moon or a planet, claiming that they are the common heritage of humankind. Article II of the Treaty states 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.” However, the state that launches a space object retains jurisdiction and control over that object; nation-states are also liable for damages caused by their space objects. This treaty raises questions about citizen claims to space assets and property, and has been the subject of international debate for years. While twenty years ago this debate was largely academic, the debate now assumes increasing urgency, with several private-sector companies perched on the edge of the space frontier, ready to launch mining missions and to claim space territory and resources as their own.

Commercialization and Exploitation

Both commercialization and exploitation of space resources currently reflect the competition of the free market and the ‘taking’ of space assets, such as mineral resources found on asteroids and other celestial bodies. In the case of space exploitation, humans take advantage of the asset of the space environment without considering the effect on the space environment. As a result of human activity, in this case space mining, what was the asset itself, the asteroid, is devalued as a result of human intervention. An underlying assumption pertaining to the space environment is that it is not fragile. Space developers have assumed that the space environment is so large and resilient that it is invulnerable to human intervention. This assumption does not take into account that space exploitation actually consumes assets found in the space environment, rendering the environment unsustainable.

Negotiating Value Propositions and Stakeholder Interests

An MIT-Draper laboratory study commissioned for NASA’s Constellation Program developed a stakeholder value analysis methodology to answer the question: “How can we architect a public enterprise that must accommodate numerous (possibly conflicting) views and ideas about how it should achieve its defined mission?” The study, conducted for NASA in 2005, identified five stakeholder interest groups: exploration, science, economic and commercial efforts, security, and the public. Each group embraced a different set of value propositions, often orthogonal to each other and conflicting. For example, scientists might believe that conservation and sustainability are of critical importance in a values discussion, while commercial companies favor the free market approach to space exploration, embracing space resource exploitation rather than preservation and conservation. While space enthusiasts and those already engaged in thinking about space exploration can readily make their voices heard, how would one include the inattentive public in a values dialogue? Are surrogates acceptable, such as organizations

expressing common social solidarity? The number and kind of inputs should we seek is a question that should be publicly discussed.

Humachines and Cyborgs

Although the relative benefits of robotic vs. human efforts in space shift as the result of changes in technology, risk, and cost, robotic spacecraft are currently cheaper, take less time to develop, and can go well beyond the solar system. Advocates of robotic exploration make the claim that humans should not be put at risk when robots can do the job better at significantly less cost. Another argument favoring the use of robots over humans is that the human presence often adds 'noise' to space research and experimentation. For example, movements of humans on the International Space Station can produce vibrations that can interfere with research experiments the field of materials science.

Counterarguments in favor of human exploration include the resilience and flexibility of humans to function in complex environments, to improvise and to respond quickly to new information and variables. The future promises to narrow this gap, however, as 'smart spacecraft' continue to evolve, with the ability to self-assess, self-diagnose, and self-repair autonomously, without human intervention.

Cognitive Prosthesis

Discussions of human exploration models tend to frame their arguments around the use of humans vs. robots to explore space. The human brain, on one hand, is able to react instantaneously to stimuli, based on all its experience and memory without any apparent logical search. On the other hand, computers have deep logical capabilities and computational skills. Scientists favor robots for planetary exploration to gather data and information without the cloudiness of human intervention. Proponents of human exploration assert that only humans have the flexibility to respond to unexpected challenges in space. Robots, the argument goes, may be limited by their profound analytic focus. Apollo 13 is often used to illustrate the need for humans in space—it took human ingenuity, quick responses to unexpected stimuli, and 'thinking outside of the box,' to bring the crew Apollo 13 home safely to Earth. Robots, the argument goes, are better at routine problem solving. Framing the discussion in terms of human vs. robots relies on robots to replicate human capabilities; in this model, robots act as surrogates for humans exploring space. Robots replace humans.

A more interesting model involves thinking about how to join humans with machines, adding a deeper dimension to what is now a relatively worn-out discussion. This approach promises to yield a more powerful and robust tool for space exploration. The question shifts from binary opposition to harmonious capability augmentation, focusing on how to integrate machines with humans in space exploration. One way to integrate humans with machine is to incorporate *cognitive prosthesis* into the space exploration model. Cognitive prosthesis involves the study of human cognition, studying the human being as a system. Based on this knowledge, the focus of this activity is to augment the capabilities of the human and overcome their limitations. The idea is not to replicate a human being through robotics, but to augment human capabilities by joining humans with machines. The use of eyeglasses to see better is an example of a prosthetic in action: the glasses augment the performance of human eyes, but do not replace them. In this way, the Curiosity Rover on the surface of Mars is also an example of using robots to augment scientists' ability to experience the surface of Mars.

Effective use of cognitive prosthesis in space exploration shifts the frame to focus on designing human/machine interaction as a system, integrating capabilities and design to create a new

performative entity—a ‘humachine’ or ‘anthroid’ whose performance merges and exploits the strengths and capabilities of human and machine. This model embodies a systems view in which human capabilities, thoughts, and actions are inextricably linked with complex technological systems, designed to amplify human cognition and perceptual abilities and performance. The result is synergistic melding of both to create a human-techno system that incorporates human and machine, extending the human’s reach and ability to see and understand the universe. Inevitably, this union performs as a kind of superhuman, endowed with the logic and analytic capabilities of machine intelligence joined with the flexibility and intuition of the human brain. The challenge, according to Kenneth Ford, director of the Institute for the Interdisciplinary Study of Human & Machine Cognition (IMHC) at the University of West Florida, is to build a human/machine system to “shift from making artificial super humans who replace us to making super-humanly intelligent artifacts that can amplify and support our own cognitive abilities” (Ford, 2004, n.p.).

The intimate blending of humans and machines via the use of cognitive prostheses ultimately yields a regime where the distinction between augmented technology-mediated humans and human mediated machines ultimately blurs. If the human species evolves beyond recognition into something else, it is not clear that there will be a need for discussions of ethics and moral reasoning.

Note

1 The opinions expressed in this paper are the author’s and do not reflect NASA positions or policies.

References

- Anderson, R. (2016). Interview With Elon Musk. *Aeon* Available at: <https://aeon.co/essays/elon-musk-puts-his-case-for-a-multi-planet-civilisation>
- Bradley, General O. N. (1967). An Armistice Day Address. In *The Collected Writings of General Omar N. Bradley*, 1, 584–589. Available at: www.opinionbug.com/2109/armistice-day-1948-address-general-omar-n-bradley/
- Cowing, K. (2016). *Changing the Way We Explore Space*, NASAWATCH April 28, 2016. Available at: <http://nasawatch.com/archives/2016/04/changing-the-wa.html#more>
- Garber, M. (2012). The Trash We’ve Left on the Moon. *Atlantic Magazine*, December 19. Available at: www.theatlantic.com/technology/archive/2012/12/the-trash-weve-left-on-the-moon/266465/
- Hildreth, S., and Arnold, A. (2014). *Threats to U.S. National Security Interest in Space: Orbital Debris Mitigation and Removal*. Washington, DC: Congressional Research Service.
- Johnson Space Center. *Orbital Debris Quarterly News*. Available at: <http://orbitaldebris.jsc.nasa.gov/newsletter/newsletter.html>.
- Kempton, W., Rayner, S., Harris, J., and Marker, A. (2001). *American Values Relating to Space and Undersea Programs*. Study for NASA Headquarters, Washington, D.C.
- McLean, M. (1998). Who Owns The Moon? *Issues in Ethics*, 10(1). Available at: www.scu.edu/SCU/Centers/Ethics/publications/iie/v101/moon.shtml
- McLean, M. (2006). *Ethical Considerations for Space Exploration*. Markkula Center for Space Ethics, Santa Clara University. Available at: www.scu.edu/ethics/focus-areas/more/resources/ethical-considerations-for-space-exploration
- MIT (2005). *The Concept Exploration & Refinement Study: Final Report, Prepared by MIT and Charles Stark Draper Laboratory for NASA’s Constellation Program*. Boston, MA.
- National Research Council (2014). *Pathways to Exploration: Rationales and Approaches for a U.S. Program of Human Space Exploration*. Washington, DC: National Academies Press. Available at: <http://nap.edu/18801>.
- PBS (1999). *Voyage to the Milky Way*. Television Show, May 19.

- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Sagan, C. (1994). *Pale Blue Dot*. New York: Random House.
- Sparrow, R. (1999). The Ethics of Terraforming. *Environmental Ethics*, 21, 227–245.
- Turner, F. (1893). *The Frontier in American History*. Available at: <http://xroads.virginia.edu/~HYPER/TURNER/> accessed 5/9/16.
- United Nations Outer Space Treaty of 1967. Available at: www.state.gov/www/global/arms/treaties/spacel.html
- Williamson, M. (2004). Protection of the Space Environment—A Time for Policy. Paper presented at the 55th International Astronautical Congress.

Further Readings

- Finney, B., and Jones, E. M. (eds.) (1985). *Interstellar Migration and the Human Experience*. Berkeley, CA: University of California Press. (A comprehensive look at social and cultural issues created by human interstellar migration.)
- Fogg, M. (1995). *Terraforming Planetary Environments*. Warrendale, PA: Society of Automotive Engineers. (The classic on terraforming from a technical point of view.)
- Geppert, A. (ed.) (2012). *Imagining Outer Space: European Astroculture in the Twentieth Century*. New York: McMillan (A thorough discussion of issues in space exploration.)
- Hargrove, E. C. (ed.) (1986). *Beyond Spaceship Earth: Environmental Ethics and the Solar System*. San Francisco, CA: Sierra Club Books. (A sourcebook on environmental ethics and outer space.)
- Lakoff, G., and Johnson, M. (1981). *Metaphors We Live By*. Chicago: University of Chicago Press. (A seminal introduction on how metaphor creates meaning and filters our understanding of the world.)
- MacNiven, D. (ed.) (1990). *Moral Expertise: Studies in Practical and Professional Ethics*. New York: Routledge. (An extensive discussion of practical ethics.)
- Rummel, J. D., Race, M. S., Horneck, G., and the Princeton Workshop Participants (2012). Ethical Considerations for Planetary Protection in Space Exploration: A Workshop, *Astrobiology*, 2, November 11. (This paper provides a key summary of issues in planetary protection by the experts.)

Transcendence in Space

Joseph C. Pitt

Introduction: Transcendence

Since the launch of Sputnik in October 1957, humanity has been space bound. But even before that, humans have yearned for the heavens. Prior to the modern era there was little we could do but gaze upward and wonder. But Sputnik opened the Space Age, and, since that event, we have been slowly, little by little, taking steps to physically explore at least our own solar system. However, aside from a few missions to the moon, the exploration of the solar system has taken place by machines. The distances involved, the extreme environmental challenges, and the toll on human psychology and physiology seem destined to restrict the human presence in space severely. Yes, we may make it to Mars and most certainly back to the moon—but beyond that, little seems feasible.

The problem is not the technologies we need actively to explore: don't measure what lies down the road by current achievements. Consider the enormous strides we have made since Sputnik: rovers on Mars and landing a probe on an asteroid! The Voyagers have begun to leave our system and make their ways into the vast beyond. Our presence will be known, just not us in person. Future technological developments will make technological exploration seem like magic—but we will be stuck here, primarily because of these too too-tender bodies we inhabit.

Many different ways of moving humans around in space have been suggested, primarily by science fiction writers. Cryogenics is a favorite means, but consider what happens when the computer controlling the ship and the life-support systems goes rogue as in the film *2001, A Space Odyssey*. It has been suggested that we build space-born ships that are actually cities (James Blish, *Cities in Flight* series), so as to provide for the time it will take to travel from A to B. Then there are the speculative advances in physics and technology such as the popular warp drive from the TV and film franchise *Star Trek*. The point is that many have recognized the difficulties of human space travel and come up with some pretty imaginative solutions. I want to propose one more and then to consider some of the problems it poses.

What if we could transcend our current physiological limitations? What if we could leave behind the bodies that limit our potential? What if we could download whatever it is that we are—dare I say our essence—into a computer on a spaceship?

Wait, don't dismiss this out of hand. Consider that we really can't predict what technological developments will be taking place in the next one hundred years. Further, we can't predict what we will come to understand about who we are and how we work and what makes us tick. Consider the strides neurophysiology has made in the last twenty years. If that kind of progress continues and even accelerates, in one hundred years we might just be able to download whatever it is that we are into whatever technological system has been developed, and escape.¹

If we think in terms of what we currently know, we are probably envisioning a space-going vehicle that has a lot of room for things like humans and robots to move around in. We will

also, on this model, need to provide for food and waste and sanitary factors. Then there is the longevity factor. We are probably thinking that if whatever we are we might be able to be transferred to some kind of robot, thereby eliminating the need for food, etc. But this not a necessary scenario. Imagine a human as a space-faring technological system with what ‘we’ are functioning as the operating system. In short, think of *us* as the spaceship itself. One day you are a typical human—mind plus body (*pace* Descartes)—and the next you are this fancy device that is capable of exploring the universe. Yes, we can’t do that today, but nothing says we can’t do it tomorrow.

So assume it is possible and let’s consider the epistemological consequences. Why the ‘epistemological consequences’? Well, the whole point of exploring the universe is to find out what is out there—to obtain knowledge, isn’t it?

Epistemology for Space

It is daunting to consider how much of Western epistemological thought is based on the physiological makeup of human beings. Think of the endless discussions of what it means for knowledge to be based on or in experience, and the extent to which it is *human* experience, based on the input of our five senses: how do we translate the buzz of the sensory in out into something coherent? Consider how much time we have spent trying to understand the connection between experience, thought, and action, when it is our actions that need explaining. Reflect on the endless attempts to define ‘thought,’ ‘thinking,’ ‘rational,’ and ‘decision making.’

I have argued elsewhere (Pitt 2016) that much of what philosophy does is refine very difficult and confused concepts to the point where they become the subject of a science. ‘Thinking’ and ‘thought’ are rapidly approaching the end of their philosophical run as brain science and computer science mature and continue to unravel the mysteries of human cognition. What will be possible once we scientifically understand how human beings go from the buzz of sensory input to coherent thinking and knowledge? Would it not be exciting to contemplate different ways to achieve those ends? That has been the goal of artificial intelligence research and it is beginning to connect with neurophysiology to probe the possibilities of neuro-networks. Once we figure out the answer to the sensory input to knowledge/action output matrix, we can start to devise other means to achieve that same end.

Thus, the human/spaceship (henceforth *humeship*) will need to have the means to find out what is out there and how to get from here to there. But are analogues of seeing, hearing, tasting, feeling, and smelling the best way to explore our environments? It may be that something like that is needed if we intend to merely recreate human beings in a different mode, capable of doing what we do in pretty much the way we do it. But as Quine noted in “Natural Kinds,” the senses we have are the result of our evolution and do not necessarily give us access to all there is—they suffice for survival on this planet.

If you adhere to the principles of evolution, there is no reason to assume the current state of human evolution is an endpoint. In fact, it would be a pretty sad commentary on the process if it were. So what we need to do is build for the future. What can/should human beings in their new form(s) evolve into? Why should we be limited by the old input model? Are there potentially new ways of gathering data that we should allow for?

Consider the paradigmatic human sense: sight. Contrary to common opinion, we do not see only with our eyes. Not only do optical telescopes and microscopes extend our visual field, but other types of telescopes and microscopes see things we can’t see with our biological optical equipment, viz., radio telescopes and electron microscopes (see Pitt 2005). Without these instruments we would not know what else is out there beyond what we can see unaided with our eyes. So, what if our new *humeship* had the means to access that kind of data as well as dark

matter, etc., directly? What if those instruments are as much a part of me the ship as my eyes are part of me the biological human? What if we have multiple counterpart senses for the other human senses? How many senses can I potentially have, and how do we integrate their inputs in something called ‘knowledge’? If we merely attempt to extrapolate from the biological human model, I suspect we will fail. The new knowledge machine, the humeship, will require a new input/output model. I predict it will require a new account of knowledge and perhaps even a new language—for what we will come to know about the universe will not be reducible to what we can discover using merely the five senses of contemporary human beings.

The Mark of Knowledge

The effort to come to some agreement on a definition of ‘knowledge’ is as old as the Western philosophical tradition. To date, for the most part, that effort has failed. The reason it has failed is that, for the most part, it has focused on the idea that knowledge is the product of a single individual who is the creator, possessor, and user of that knowledge. So, the problem has been to figure out a solution to the input/output problem noted earlier—but that just may be the wrong focus.

Individuals do not create knowledge, communities do. At least you need a community to take what an individual proposes and grant it the status of knowledge. This was the insight of C. S. Peirce and a fundamental feature of pragmatism. Communities of experts, as Peirce saw it, formulate the criteria by which a proposal is granted the status of knowledge. Those criteria can and do change over time, as it is deemed adequate or not to the job. This also means that the knowledge itself changes, as what on an early set of criteria was considered knowledge turns out not to count on a new set of criteria. To a certain degree we already know this since we know that the history of science is the history of failed theories. So it looks like the ‘true for ever’ view of knowledge won’t work. That doesn’t mean we have to give up the truth condition. It just means we need to weaken it to ‘true for now.’ Peirce was unwilling to give up the strong condition. For Peirce, the strong condition for truth was reality:

The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real. That is the way I would explain reality. . . . Our perversity and that of others may indefinitely postpone the settlement of opinion; it may even conceivably cause an arbitrary proposition to be universally accepted as long as the human race should last. Yet even that would not change the nature of the belief, which alone could be the result of investigation carried sufficiently far; and if after the extinction of our race, another should arise with faculties and disposition for investigation, that true opinion must be the one which they would ultimately come to.
(Peirce 1878, 1955: 38–39)

For a pragmatist, the mark of knowledge is successful action. As C. I. Lewis puts it:

Knowledge, action, and evaluation are essentially connected. The primary and pervasive significance of knowledge is for the sake of doing. And action, obviously is routed in evaluation. For a being which did not assign comparative values, deliberate action would be pointless; and for one which did not know, it would be impossible.

(Lewis 1946: 3)

Successful action is determined if, when acting on community-endorsed knowledge, one achieves one’s goals. Whatever it may mean, successful action requires (among other things) the ability to take into account all the relevant data, a coherent set of goals, and a set of values.

The process of generating knowledge is iterative, adjusting what we think we know and our goals and values in the light of what we try to do.²

The epistemological issues here are many and varied. If the projected humeship is indeed equipped with many more sources of data input means than our current five, integrating them into a coherent whole will require some very sophisticated programming (used in the broadest possible sense, *pace* Sellars). Right now we haven't figured out how the human brain coordinates inputs from our five senses plus our reasoning faculty. Anticipating constructing a system that has multiples of our five senses without knowing what they will be will require thinking outside the box (so to speak). The system we construct will not be complete, so it must have the means to reprogram itself. (I am speaking in computer programming terms simply because that is the only model available at the moment that makes sense.) I don't think this is an impossible barrier, for we seem to be doing just that now. Consider how the manner in which we seek information has changed and how this has happened in a very short time period. I am referring to the role smartphones now play in our lives. If we want to know something, from the date of Galileo's birth to last night's football score to the location of a good restaurant, just 'google it.' This has become an almost automatic reaction to a 'what' question and it is not just the kids who are reacting this way.

So we can and already do adjust to new ways of accessing data. The question is, will we be able to construct a mostly artificial system that can do this? It seems we must if we are to construct my humeship.

So far we have been considering factors that the technologies for the humeship raise. Is there something particularly human we need to ensure for? To answer that I suspect we need to address that old chestnut: "What makes us human?" Clearly that is too large an issue to be resolved in this short chapter. But we can point to several features of our species that seem salient: curiosity (but other species, e.g., dogs, cats, are curious); the ability to reason; a sense of humor; the capacity for love, passion, and compassion; the ability to learn from our mistakes; and the ability to make and keep a promise. Of course, this is an incomplete list—but the point is to ensure that those qualities that make us what we are also part of the humeship. But there are other qualities worth considering, if only to consider not including them. Human beings can be cruel. We are also warlike. We can lie. We are capable of betrayal.

But what we are is perhaps most clearly captured by what has been called 'the human condition'—we are the only species, as far as we can tell, whose individual members know they are going to die. So, if we transform ourselves into a self-repairing, self-reprogramming entity that is, for all practical purposes, immortal, how does that change our understanding of what it is to be human?

Now it might be argued that since the starting premise is that we 'download our essence,' are we talking about the equivalent of DNA editing if we want to not include our less-admirable qualities? Do we know enough about how the negatives influence the positives and *visa versa*? That is, can we love if we can't be cruel? What does it mean to keep a promise if we can't lie? If we are immortal, does that mean we need to pay more attention to whatever relationships we have?

I do not have the answers to these questions, but I am convinced they are important. Equally important is that we figure out a way to get off this planet before we destroy it and ourselves.

Talking about saving ourselves raises another issue, really complicating the problem: Why do we have to think about the humeship as containing only one download? But if there are many, how do we ensure peace and cooperation? What are the selection criteria for who is to be chosen? Must those who are downloaded have extensive technical capabilities or can we hand that off to some other part of the humeship? Would multiple downloads open up the possibility of a hive mind? Multiple downloads would solve the problem of loneliness, but at what price?

There is another consideration we must address when we think of what is involved in a hive mind: loss of mental flexibility. One of the less-understood features of the growth of knowledge

is the dynamic involving the individual and the group. Each of us is stimulated through our senses. Using whatever program we operate, we transform that experiential buzz into propositions by which we try to capture the nature of our experience and the world. Not only do we refine the end result as we increasingly interact with the world, but in interacting with our fellow knowers we refine the sum total of those propositions to produce a consensus we call 'knowledge.' Several things need to be noted here. Each of us processes our inputs differently. I would argue that our language is our processing program. But even though we manage to communicate somewhat with each other, the number of misunderstandings and disagreements about what we can know suggests that we learn and use that program embedded in the language we share differently. That is a good thing—for it is a major source of creativity. But when we look at our ship, we have to consider what kind of 'senses' we will build into it. As Quine suggests in "Natural Kinds," the senses we have are largely a function of having served us well in the race for survival here on Earth (Quine 1969: 126). From this it does not follow that our senses give us access to the world *as it really is*. So, to survive in the universe won't we need to equip our humeship with the senses to accommodate more than what we experience on Earth? If not that, we need at least to give it the ability to modify/extend/add/change the original senses we give it to deal with the universe it discovers. When we look at the components of the hive mind, if they are a bunch of us, then they will not form an 'organic unity' at first. So what will happen? It is possible that over time such a unity may develop—but is that a good thing? Will we not lose our flexibility to respond to whatever new and unexpected things we encounter, since we now will have only one way to think about things? Being able to think about things from a variety of viewpoints is an essential part of being human. Will this mean the end of humanity as we know it?

The hive mind isn't the only option. Having enough computing power to safely react to the universe we discover may mean that each person can be downloaded intact and we have a cyber community in our humeship. That raises the problem of the organization of community. Someone needs to make decisions. Democracy does not seem to work too well on Earth: would it fare better in a cyber state? But given the extreme technical demands, wouldn't a meritocracy/technocracy provide better leadership? Interesting problem.

Finally, human beings can be myopic at times. When we talk about going into space, we tend to forget that we have evolved on this planet in tandem with a host of species, some of which are essential to our mental well-being. Dogs, horses, cats all contribute to making us human. But how are we to take them with us? Even if we somehow figure out how to isolate the 'essence' of this dog or that cat, making them part of a cyber society makes little sense—dogs live their lives with a heightened sense of smell, cats need a sunbeam to curl up in, horses need to run. Living with these species has become part of our very being. One of my favorite sayings from the world of horses is "there is nothing as good for the inside of a man or woman than the outside of a horse." However, if we leave them behind, we leave part of ourselves and our evolutionary history behind—and then we will truly be transcending the Earth-bound human.

It seems at this point we can only venture into further speculation. The purpose of this chapter is to get us to think about a serious problem that has serious consequences for us as a species and for us as *human* beings. We will need to leave the planet at some point. How are we going to do it, and in so doing what will that do to us?

Notes

- 1 Of course, that assumes the planet will be able to sustain human life for a long-enough period for this to develop.
- 2 For more fully developed account of this take on 'knowledge,' see Pitt 2000.

References

- James Blish (1950–1962), Cities in Flight series.
- Lewis, C. I. (1946). *An Analysis of knowledge and Valuation*. LaSalle, IL: Open Court.
- Pitt, J. C. (2000). *Thinking About Technology*. Available at: www.phil.vt.edu/Pitt/pittbook_2.pdf
- Pitt, J. C. (2005). When Is an Image not an Image? *Techné* (8)3.
- Pitt, J. C. (2016). The Future of Philosophy; a Manifesto. In P. Kroes and P. Vermass (eds.) *The Empirical Turn in the Philosophy of Technology Revisited*. Amsterdam: Springer.
- Quine, W. V. O (1969). Natural Kinds. In *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- Peirce, C. S. (1878, 1955). How to Make Our Ideas Clear. In J. Buchler (ed.) *Philosophical Writings of Peirce*. New York: Dover Publications.

The Role of Technology in Natural Science

Nicholas Rescher

The Empirical Impetus

Modern science is critically dependent upon technology, and scientific progress almost invariably demands an escalation in the technological state of the art—for natural science is fundamentally empirical, with its advance dependent not on human ingenuity alone but on the monitoring observations to which we can gain access only through interactions with nature. The days are long past when useful scientific data can be had by unaided sensory observation of the ordinary course of nature. Artifice has become an indispensable route to the acquisition and processing of scientifically useful data. The sorts of data on which scientific discovery nowadays depends can be generated by technological means only.

With the enhancement of scientific technology, the size and complexity of this body of data inevitably grows, expanding on quantity and diversifying in kind. Technological progress constantly enlarges the window through which we look upon nature's parametric space. In developing natural science, we continually enlarge our view of this space and then generalize upon what we see. But what we have here is not a homogeneous lunar landscape, where once we have seen one sector we have seen it all, and where theory projections from lesser data generally remain in place when further data come our way. Instead, historical experience shows that there is every reason to expect that our ideas about nature are subject to constant radical changes as we explore parametric space more extensively. The technologically mediated entry into new regions of parametric space constantly destabilizes the attained equilibrium between data and theory. Technology is the engine that advances the frontier of progress in natural science.

The idea of 'novel phenomena' is the talisman of contemporary experimental physics. Workers in this domain do not nowadays talk of increased precision or accuracy of measurement in their own right. It is all a matter of the search for new phenomena—and the development of new data-technology is virtually mandatory for the discerning of new phenomena. The reason is simple enough: inquiry proceeds on such a broad front nowadays that the range of phenomena to which the old data-technology gives access is soon exhausted. The same is generally true with respect to the range of theory testing and confirmation that the old phenomena can underwrite.¹

When such limitations in the database occur, they produce correlative imperfections in our theoretical understanding. A homely fishing analogy of A. S. Eddington's is useful here.² He saw the experimentalist as one who trawls nature with the "net" of his equipment for detection and observation. Now suppose (says Eddington) that fisherman trawls the seas using a fishing net of 2-inch mesh. Then fish of a smaller size will simply go uncaught—and the theorists who analyze the experimentalist's catch will have an incomplete and distorted view of aquatic life. Only by improving our observational means for 'trawling' nature can such imperfections be mitigated.

Physicists often remark that the development of our understanding of nature moves through successive layers of *theoretical* sophistication.³ But scientific progress is clearly no less dependent on continual improvements in strictly *technical* sophistication:

Some of the most startling technological advances in our time are closely associated with basic research. As compared with 25 years ago, the highest vacuum readily achievable has improved more than a thousand-fold; materials can be manufactured that are 100 times purer; the submicroscopic world can be seen at 10 time higher magnification; the detection of trace impurities is hundreds of times more sensitive; the identification of molecular species (as in various forms of chromatography) is immeasurably advanced. These examples are only a small sample. . . . Fundamental research in physics is crucially dependent on advanced technology, and is becoming more so.

(National Research Council [U.S.] Physics Survey Committee 1973: 23)

Without an ever-developing technology of experimentation and observation, scientific progress would grind to a halt. The discoveries of today cannot be attained with yesterday's instrumentation and techniques. To secure new observations, to test new hypotheses, and to detect new phenomena, an ever more powerful technology of inquiry is needed. Throughout the natural sciences, technological progress is a crucial requisite for cognitive progress. We are embarked on an endless endeavor to improve the range of effective observational and experimental intervention. Only by operating under new and previously inaccessible conditions—attaining extreme temperature, pressure, particle velocity, field strength, and so on—can we realize those circumstances that enable us to put our hypotheses and theories to the test. As an acute observer has rightly remarked: “Most critical experiments [in physics] planned today, if they had to be constrained within the technology of even ten years ago, would be seriously compromised” (National Research Council [U.S.] Physics Survey Committee 1973: 16).

Science has accordingly come to outstrip rivaling military weaponry industry in its concern for technological sophistication:

[F]undamental research in physics is crucially dependent on advanced technology, and is becoming more so. Historical examples are overwhelmingly numerous. The postwar resurgence in low-temperature physics depended on the commercial production of the Collins liquefier, a technological achievement that also helped to launch an era of cryogenic engineering. And today, superconducting magnets for a giant bubble chamber are available only because of the strenuous industrial effort that followed the discovery of hard superconductors. In experimental nuclear physics, high-energy physics, and astronomy—in fact, wherever photons are counted, which includes much of fundamental physics—photomultiplier technology has often paced experimental progress. The multidirectional impact of semiconductor technology on experimental physics is obvious. In several branches of fundamental physics it extends from the particle detector through nanosecond circuitry to the computer output of analyzed data. Most critical experiments planned today, if they had to be constrained within the technology of even ten years ago, would be seriously compromised.

(National Research Council [U.S.] Physics Survey Committee 1973: 23)

Recognition of reciprocal dependence between scientific and technological progress does not, however, affect the basic point of the present deliberations. To be sure, technology consists largely of applied science, and technological innovation in some degree reflects new uses of scientific knowledge. But this symbiosis does not displace the crucial consideration that

science cannot effectively push its inquiry into nature further than the available mechanisms of informational technology permit. These limits are not only limitations on one's capacity to test hypotheses that are already in hand (as the ancients could not test hypotheses about mountains on the far side of the moon), but—more crucially—they impose restrictions on one's capacity even to conceptualize certain hypotheses (as the ancients could have no glimmering of the red shift). Progress is insuperably limited at any given stage of scientific history by the implicit barriers set by the available technology of data acquisition and processing.

Is Completion a Prospect? The Technological Limits and Limitations of Inquiry

Technology extends the capacity of our natural economy in point of observation with nature. And the arbitrament of praxis—not theoretical merit but practical capability—affords the best standard of adequacy for our scientific proceedings that is available. But could we ever be in a position to claim that science has been completed on the basis of the success of its practical applications? On this basis, the perfection of science would have to manifest itself in the perfecting of control—in achieving a perfected technology. But just how are we to proceed here? Could our natural science achieve manifest perfection on the side of control over nature? Could it ever underwrite a recognizably perfected technology?

The issue of 'control over nature' involves much more complexity than may appear on first view. For just how is this conception to be understood? Clearly, in terms of bending the course of events to our will, of attaining our ends within nature. But this involvement of *our ends* brings to light the prominence of our own contribution. For example, if we are inordinately modest in our demands (or very unimaginative), we may even achieve 'complete control over nature' in the sense of being in a position to do *whatever we want* to do, but yet attain this happy condition in a way that betokens very little real capability.

One might, to be sure, involve the idea of omnipotence and construe a 'perfected' technology as one that would enable us to do literally *anything*. But this approach would at once run into the old difficulties already familiar to the medieval scholastics. They were faced with the challenge: "If God is omnipotent, can he annihilate himself (contra his nature as a *necessary* being), or can he do evil deeds (contra his nature as a *perfect* being), or can he make triangles have four angles (contrary to *their* definitive nature)?" Sensibly enough, the scholastics inclined to solve these difficulties by maintaining that an omnipotent God need not be in a position to do literally anything but rather simply anything that it *is possible* for him to do. Similarly, we cannot explicate the idea of technological omnipotence in terms of a capacity to produce and result, wholly without qualification. We cannot ask for the production of a *perpetuum mobile*, for spaceships with 'hyperdrive' enabling them to attain transluminal velocities, for devices that predict essentially stochastic processes such as the disintegrations of transuranic atoms, or for piston devices that enable us to set *independently* the values for the pressure, temperature, and volume of a body of gas. We cannot, in sum, ask of a 'perfected' technology that it should enable us to do anything that we might take it into our heads to do, no matter how 'unrealistic' this might be.

All that we can reasonably ask is that perfected technology should enable us to do anything *that it is possible for us to do*—and not just what we might *think* we can do but what we really and truly can do. A perfected technology would be one that enabled us to do anything that *can possibly* be done by creatures circumstanced as we are. But how can we deal with the pivotal conception of 'can' that is at issue here? Clearly, only science—real, true, correct, *perfected* science—could tell us what indeed is realistically possible and what circumstances are indeed inescapable. Whenever our 'knowledge' falls short of this, we may well 'ask the impossible' by

way of accomplishment (for example, spaceships in ‘hyperdrive’), and thus complain of incapacity to achieve control in ways that put unfair burdens on this conception.

Power is a matter of the ‘effecting of things possible’—of achieving control—and it is clearly cognitive state-of-the-art in science that, in teaching us about the limits of the possible, is itself the agent that must shape our conception of this issue. *Every* law of nature serves to set the boundary between what is genuinely possible and what is not, between what can be done and what cannot, between which questions we can properly ask and which we cannot. We cannot satisfactorily monitor the adequacy and completeness of our science by its ability to effect ‘all things possible,’ because science alone can inform us about what is possible. As science grows and develops, it poses new issues of power and control, reformulating and reshaping those demands whose realization represents ‘control over nature.’ For science itself brings new possibilities to light. (At a suitable stage, the idea of ‘splitting the atom’ will no longer seem a contradiction in terms.) To see if a given state of technology meets the condition of perfection, we must *already* have a body of perfected science in hand to tell us what is indeed possible. To validate the claim that our technology is perfected, we need to *preestablish* the completeness of our science. The idea works in such a way that claims to perfected control can rest only on perfected science.

In attempting to travel the practicalist route to cognitive completeness, we are thus trapped in a circle. Short of having supposedly perfected science in hand, we could not say what a perfected technology would be like, and thus we could not possibly monitor the perfection of science in terms of the technology that it underwrites.

Moreover, even if (*per impossible*) a ‘pragmatic equilibrium’ between what we can and what we wish to do in science were to be realized, we could not be warrantedly confident that this condition will remain unchanged. The possibility that ‘just around the corner things will become unstuck’ can never be eliminated. Even if we ‘achieve control’ for all visible intents and purposes, we cannot be sure of not losing our grip upon it: not because of a loss of power but because of cognitive changes that produce a broadening of the imagination and a widened apprehension as to what ‘having control’ involves—for the project of achieving practical mastery can never be perfected in a satisfactory way. The point is that control hinges on what we want, and what we want is conditioned by what we think possible, and *this* is something that hinges crucially on theory—on our beliefs about how things work in this world. So control is something deeply theory-infected. We can never safely move from apparent to real adequacy in this regard. We cannot adequately assure that seeming perfection is more than just that. We thus have no alternative but to *presume* that our knowledge (that is, our purported knowledge) is inadequate at this and indeed at any other particular stage of the game of cognitive completeness.

Such a perspective indicates that our knowledge of nature is limited in two dimensions. There are both thematic and progressive limits. The first is a matter of the specific *parametric direction* in which our inquiries orient our data-seeking interactions with physical reality. The second is a matter of the extent to which we are willing and able to push forward in this direction.

The cognitive yield of such a project is not continuous but proceeds stepwise: new data disestablishes old theories whose revision in turn provides a new impetus to data acquisition. The result is a dialectic of to and fro as between theory systematization and technological advance.

The Potential of Alien Science

To what extent would the *functional equivalent* of our physical science built up by the inquiring intelligences of an astronomically remote civilization be bound to resemble our science?

In reflecting on this question and its ramifications, one soon comes to realize that there is an enormous potential for diversity.

Williams James wrote:

Were we lobsters, or bees, it might be that our organization would have led to our using quite different modes from these [actual ones] of apprehending our experiences. It *might* be too (we cannot dogmatically deny this) that such categories unimaginable by us to-day, would have proved on the whole as serviceable for handling our experiences mentally as those we actually use.

(1907: 171)

One's language and thought processes are bound to be closely geared to the world as one experiences it. As is illustrated by the difficulties we ourselves experience in bringing the language of everyday experience to bear on subatomic phenomena, our concepts are ill-attuned to facets of nature different in scale or structure from our own. We can hardly expect a 'science' that reflects such parochial preoccupations to be a universal fixture.

With alien science viewed in this perspective, we see that the *machinery of formulation* used in expressing their science might be altogether different. Specifically, alien mathematics might be very unlike ours. Their dealings with quantity might be entirely anumerical—purely comparative, for example, rather than quantitative. Especially if their environment is not amply endowed with solid objects or stable structures congenial to measurement—if, for example, they were jellyfish-like creatures swimming about in a soupy sea—their 'geometry' could be something rather strange, largely topological, say, and geared to flexible structures rather than fixed sizes or shapes. Digital thinking might be undeveloped, while certain sorts of analogue reasoning might be highly refined. Or, again, an alien civilization might, like the ancient Greeks, have 'Euclidean' geometry without analysis. In any case, given that the mathematical mechanisms at their disposal could be very different from ours, it is clear that their description of nature in mathematical terms could also be very different. (Not necessarily truer or falsier, but just different.)

A comparison of the 'science' of different civilizations here on Earth suggests that it is not an outlandish hypothesis to suppose that the very *topics* of alien science might differ dramatically from those of ours. In our own case, for example, the fact that we live on the surface of the Earth (unlike whales), the fact that we have eyes (unlike worms) and thus can *see* the heavens, the fact that we are so situated that the seasonal positions of heavenly bodies are intricately connected with agriculture—all these facts are clearly connected with the development of astronomy. The fact that distant creatures would experience nature in ways very different from ourselves means that they can be expected to raise very different sorts of questions. Indeed, the mode of emplacement within nature of alien inquirers might be so different as to focus their attention on entirely different aspects of constituents of the cosmos. If the world is sufficiently complex and multifaceted, they might concentrate upon aspects of their environment that mean nothing to us, with the result that their natural science is oriented in directions very different from ours.⁴

To motivate this idea of a conceptually different science, it helps to cast the issue in temporal rather than spatial terms. The descriptive characterization of *alien* science is a project rather akin in its difficulty to that of describing our own *future* science. It is a key fact of life that progress in science is a process of *ideational* innovation that always places certain developments outside the intellectual horizons of earlier workers. The very concepts we think in terms of become available only in the course of scientific discovery itself. Like the science of the remote future, the science of remote aliens must be presumed to be such that we really could not achieve intellectual access to it on the basis of our own position in the cognitive scheme of

things. Just as the technology of a more advanced civilization would be bound to strike us as magic, so its science would be bound to strike us as incomprehensible gibberish—until we had learned it ‘from the ground up.’ They might (just barely) be able to *teach* it to us, but they could not *explain* it to us by transposing it into our terms.

It is tempting to reason: “Since there is only one nature, only one science of nature is possible.” Yet, on closer scrutiny, this reasoning becomes highly problematic.

Physical laws are pervasive regularities in nature and are always and everywhere the same. But detection will of course vary drastically with the mode of observations—that is, with the sort of resources that different creatures have at their disposal to do their detecting. Everything depends on how nature pushes back on our senses and their instrumental extensions. Even if we detect everything we can, we will not have got hold of everything available to others. The converse is equally true. The laws that we (or anybody else) manage to formulate will depend crucially on one’s place within nature—on how one is connected into its wiring diagram, so to speak.

Empirical science is always the result of *inquiry* into nature, and this is inevitably a matter of a *transaction* or *interaction* in which nature is but one party and the inquiring beings another. We must expect alien beings to question nature in ways very different from our own. On the basis of an *interactionist* model, there is no reason to think that the sciences of different civilizations will exhibit anything more than the roughest sorts of family resemblance.

Admittedly, there is only one universe, and its laws and materials are—far as we can tell—the same everywhere. We share this common universe with all life-forms. However radically we differ in other respects (in particular, those relating to environment, to natural endowments, and to style or civilization), we have a common background of cosmic evolution and a common heritage of natural laws. So, if intelligent aliens investigate nature at all, they will investigate the same nature we ourselves do. All this can be agreed. But the fact remains that the corpus of scientific information—ours or anyone’s—is an ideational construction, and the sameness of the object of contemplation does nothing to guarantee the sameness of ideas about it.

So, in the end, our human sort of physical science may well be *sui generis*, adjusted to and coordinated with a being of our physical constitution, inserted into the orbit of the world’s processes and history in our sort of way. It seems that alike in science and technology, as in other areas of human endeavor, we are prisoners of the thought-world that our biological and social and intellectual heritage affords us.

Notes

- 1 Use of the term *phenomena* and insistence upon their primacy in scientific inquiry goes back to the ancient idea of “saving the phenomena” (see Pierre Duhem, *To Save the Phenomena*, trans. E. Doland and C. Maschler [Chicago, 1969], and Jürgen Mittelstrass, *Die Rettung der Phanomene* [Berlin, 1962]). It finds its modern articulation in Newton’s *Principia*. The key point presently at issue was clearly perceived by Karl Pearson at the turn of the century:

The progress of science lies in the continual discovery of more and more comprehensive [i.e., *comprehensively applicable*] formulae by and of which we can clarify the relationships and sequences of *more and more extensive groups of phenomena*. (*The Grammar of Science*, ed. E. Nagel [New York, 1957], 96–97; italics supplied.)

- 2 See A. S. Eddington, *The Nature of the Physical World* (New York, 1928).
- 3 “Looking back, one has the impression that the historical development of the physical description of the world consists of a succession of layers of knowledge of increasing generality and greater depth. Each layer has a well-defined field of validity; one has to pass beyond the limits of each to get to the next one, which will be characterized by more general and more encompassing laws and by discoveries constituting a deeper penetration into the structure of the Universe than the layers recognized before.” See Edoardo Amaldi, “The Unity of Physics,” *Physics Today*, vol. 261, no. 9 (September 1973), 18–32 and also E. P. Wigner, “The Unreasonable Effectiveness of Mathematics in the Natural Sciences,”

Communication on Pure and Applied Mathematics, vol. 13 (1960), 1–14, as well as Wigner’s “The Limits of Science,” *Proceedings of the American Philosophical Society*, vol. 93 (1949), 521–526. Compare also Chapter 8 of Henry Margenau, *The Nature of Physical Reality* (New York: McGraw-Hill, 1950).

- 4 His anthropological investigations pointed Benjamin Lee Whorf in much this same direction. He wrote: “The real question is: What do different languages do, not with artificially isolated objects, but with the flowing face of nature in its motion, color, and changing form; with clouds, beaches, and yonder flight of birds? For as goes our segmentation of the face of nature, so goes our physics of the cosmos” (“Language and Logic,” in *Language, Thought, and Reality*, ed. by J. B. Carroll [Cambridge, MA: MIT Press, 1956], 240–241). Compare also the interesting discussion in Thomas Nagel, “What Is It Like to Be a Bat?” in *Mortal Questions* (Cambridge, MA: Harvard University Press, 1976).

References

- Amaldi, E. (1973). The Unity of Physics. *Physics Today*, 261(9), 18–32.
- National Research Council (U.S.). Physics Survey Committee (1973). *Physics in Perspective*, Student edition. Washington, DC: National Research Council/National Academy of Science Publications.
- Duhem, P. (1969). *To Save the Phenomena*, trans. E. Doland and C. Maschler. Chicago: University of Chicago Press.
- Eddington, A. S. (1928). *The Nature of the Physical World*. New York: The Macmillan Company.
- James, W. (1907). *Pragmatism*. New York: Longmans.
- Mittelstrass, J. (1962). *Die Tettubng der Phanomene*. Berlin: Walter de Gruyter GmbH & Co..
- Nagel, E. (ed.) (1957). *The Grammar of Science*. New York: Meridian Books.
- Nagel, T. (1976). What Is It Like to Be a Bat? In *Mortal Questions*. Cambridge, MA: Harvard University Press.
- Whorf, B. L. (1956). Language and Logic. In J. B. Carroll (ed.) *Language, Thought, and Reality*. Cambridge, MA: Technology Press of Massachusetts Institute of Technology and John Wiley & Sons, Inc., 240–241.
- Wigner, E. P. (1949). The Limits of Science. *Proceedings of the American Philosophical Society*, 93, 521–526.
- Wigner, E. P. (1960). The Unreasonable Effectiveness of Mathematics in the Natural Sciences. *Communication on Pure and Applied Mathematics*, 13, 1–14.

Index

- abduction, instances 200
Abrahamic religions, transcendence/
 techno-lifeworld 93–4
Absolute Spirit, Hegelian realization 154
Acoustic Technics (Ihde) 54–5
actors, notion 105
Advanced Manufacturing Partnership 217
advanced reproductive technology (ART) 226–7
Advanced Research Projects Agency
 (ARPANET) 145
aesthetic pleasure, transcendence 95
African technoart 56–7
agency: citizen agency 99; impact 98; power, role
 98–9
AirBnb service 26
airline industry, energy demand estimates 269–70
Alberti, Leon 58
aletheuein 224
algorithmic culture 47–9
algorithms, usage 169
alienation, concept 101
alien science, potential 349–51
alterity 176; games form 176–7; relations
 23–4, 188
Amazon: online recommender system 4; Web
 Services 158
Amdahl's Law 242
American Enterprise Institute (AEI) 267
Analog Days (Pinch/Trocco) 60
animality 109
Animal Liberation (Singer) 109
animals 108; agenda 113–14; anthropomorphizing
 111; cognition, understanding (absence) 112;
 humans, understanding 111–12; objections
 110–13; research 108, 113; technology concept,
 absence 112–13; themes/motivations 109–10
“Anonymity Versus Commitment” (Dreyfus) 84
anonymization, usage 167
Anstey, Joseph 67
anthropocene, technology (usage) 300–1
antiskid devices, design case study 211–13
Apple, product innovations 203–4
application exploits 143
Arab Spring 47, 162
Arendt, Hannah 252, 260–1
Argentine tango, focus 221–2
Aristophanes 154
armed robots: argument 76; autonomy,
 relationship 74–6; existential risk, relationship
 78–9; Just War Theory, relationship 76–7;
 nonlethal/defensive usage 77–8
armed robots, usage 73
Arms, Suzanne 226–7
art: humanity feature 63; nature 64;
 philosophy 62
Art, Technology and Science (Smith) 62
artificial intelligence (AI) 73, 243
artificiality 174–5
artificially intelligent armed robots,
 responsibility 73
artificial moral agents (AMAs): debate 75;
 qualification 74
artists: impact 68; technology, relationships 62;
 thought translation 66
Asimov, Isaac 78
Assman, Jan 94
Asteroid Redirect Mission (ARM) 317–19;
 downselection 319–20; goals, participant
 average priority rankings 321; goals, results
 321–2; options 319–20; options, results
 320–1
Asteroid Robotic Redirection Vehicle
 (ARRV) 319
asymmetrical kinetic warfare 149
attribution, difficulty 149
authority, decline 101–2
authorship 133–6, 220–1
Automatic Sequence Controlled Calculator
 (IBM) 209
autonomous lethal robots 73
autonomy, armed robots (relationship) 74–6
Axial Age 93, 96

background relations 24–5, 176, 188
Bacon, Francis 100
Band-Aid, ontology 14–15
Barney, Darin 177
Barthes, Roland 63

- Basset, Caroline 4
 Battista, David 269
Battle Raper: Hyper Realaction 29
 Baudrillard, Jean 44
 Beckoff, Mark 109
being-in-the-world-ness 62
 Bel Geddes, Norman 209
 Bell, Alexander Graham 202
 Bell, Daniel 245
 Belu, Dana S. 224
 Benford, Gregory 267
 Benjamin, Walter 216, 220
 Benkler, Yochai 132
 Berne, Rosalyn W. 305
 Besmer, Kirk 189
Between Reason and Experience (Feenberg) 273
 Beuys, Joseph 66
 Bhabha, Homi 220
 bias, self-fulfilling cycle 172
 Bickel, J.E. 274
 bidirectional flow 65
 Big Bang 111
 big data: advances 123; defining 164–5; ethical issues 164; expertise/promise forms 4; fairness 169–72; harms 168; mythology, confrontation 4–5; prediction 169–72; volume/velocity/variation 165
 big gods, existence 93–4
 Big Technology 113
 Bijker, Wiebe 105, 200, 202
 binoculars, usage 191
 biochar 264
 biology, philosophy (emergence) 110
 biomedically constructed birth 227–8
 birth *see* childbirth
 Bivens, Rena 10
 Bjorkman, Jonas 193
 Black Lives Matter 162
 Blockhead account, difficulties 135
 Blockhead model 130–2, 134
 Blockhead theory 133–4
 Boellstorff, Tom 5
Boom! Boom! Boom! Boom! 34
 Borgmann, Albert 44, 82–5, 89–90, 175, 177
 Bostrom, Nick 252
 Boswell, James 130–1
 bounded rationality 123
 Bowker, Geoffrey 4, 5
 boyd, danah 20
 Boyle, James 135–8
 Brey, Philip 35, 64
 Briggie, Adam 281
 Brin, David 306
 Brin, Sergey 6
 Brock, David 240
 Bucciarelli, Louis 199–200, 210–11
 Buchla, Don 60
 Buddhism 92, 95
 bureaucracy (Weber) 103
Bursac v. Suozzi 125
 Caldeira, Ken 269
 CalPERS, pension information 122
 Campbell, Joseph 63
 Campus, Peter 64
 capitalism, expansionary logic 254
 carbon dioxide removal (CDR) 264, 272; techniques 275–6
Carmageddon 30
 Cartesian dualism 45
 catastrophic economic consequences 276
ceteris paribus: assumptions 270; logic 263, 267–8, 272, 275
 Chafe, Bill 55
 Charles, Ray 135
 childbirth: biomedically constructed birth 227–8; harnessing, technology (usage) 224; interlocking paths 230–1; labor, disembodied experience 229; midwives, impact 227; nontechnological childbirth 224; pain 231–3; Smellie, impact 227; technological childbirth 224; technophilic childbirth 226–31; technophobic childbirth 231–3
Childbirth Without Fear (Dick-Reed) 231
 Christianity 94
 Cicero 219
Circle, The (Eggers) 42, 47
 cisgender 8
 cis woman, term 22
 citizen agency 99
 citizenship: impact 98; technical citizenship 98–100
 C-K theory 204
 Clark, Andy 64
 Clarke, Arthur C. 307
 Clausewitz, Carl von 148–9
 climate change 263; discourse 292–6; energy consumption, good life (relationship) 297–9; footprint minimization 296–7; IPCC perspective 292–3; political discourse 293–5; remediation 242; social/technical criteria 276–7; technological discourse 295–6; technology, philosophy 292; technology-informed approaches, philosophy 296–300
 climate justice 294
 Cloud 154; characteristics (NIST) 156–7; chattering 160; computing 160, 243; concerns 158–62; control 162; imagination 155–6; possibilities 156; reality 156–8
Clouds (Aristophanes) 154
 cloud seeding 275
 Cockburn, Cynthia 6
 code, usage 126–8
 Coeckelbergh, Mark 36
 cognitive capitalism 47

- cognitive prosthesis 337–8
 collaboration, occurrence 211
 Collins, Allan 82
 command society 47
 communication network attacks 145–6
 communicative rationality 103
 completion, prospect 348–9
 computational causal chains, usage 150
 computer-based art 66
 computer-generated output, impact 178
 computers 83–5
 Condella, Craig A. 154
 configuration exploits 143
 Confucianism 92, 94
 connectivity, culture 45
 constructivist technology, studies 105
 contemporary body 67
 content creators, payment 161
 contextual integrity 122
 Cook, Jill 55, 57
 Copenhagen Consensus Center (CCC) 274
 Copernicus 252
 copyright 130; law, understanding 131
 co-shaping 186
 cosmology 259
 Cox, Laverne 3
creation ex nihilo 63
 creativity: blockhead model 130–3; economic
 account, disruption 134
 creativity, economic/cultural models
 (copyright) 130
 Crenshaw, Kimberlé 9–10
Critical Theory of Technology
 (Feenberg) 47
 cross-context correlations 169
Crossing the Postmodern Divide (Borgmann)
 82, 83
 Crutzen, Paul 269
 cryptopticon 168
 cues, filtered-out approach 178
 cultural activity 134
 cultural relations 134–5
 culture, economy (contrast) 136–8
Custer's Revenge 29, 37
 cyberattack: action 147–9; application exploits
 143; communication network, usage 145–6;
 configuration exploits 143; defining 142–7;
 definition 148; executive system exploits
 143–4; noncommunicative interaction chains,
 usage 146; NSC definition 148; psychological
 influence 146; resource attacks 143; strategic
 influence 147; tactical influence 147;
 vulnerability 144
 cyberattack, influence: aims 146–7; means
 144–6; mode 143–4
 cyber-espionage 146
 cyber-operation 148
 cyberspace 148
 cyberwarfare 141, 149–50; capabilities, building
 150; definition 149; fear 142; trend 150–1
 cyborgs 337–8; relations 186, 189
 “Daily We,” idea 24
 dance 216; study, translation (cultural approach)
 221–2
 Daoism 92
 Darrow, Robert 251
 dashboard infotainment 192
das Man 16
 data 3; analysis 166, 171; big data, mythology
 (confrontation) 4–5; transgender lives,
 relationship 7–10
 da Vinci, Leonardo 58, 208
 Davison, Megan 8
 deadweight loss 138
Death Race 30
 decisional privacy 123
 Deep Blue 244
 Deitch, Jeffrey 67–8
 democratic citizens, activities 103
 democratic input 315
 democratic intervention, concept 104
 democratic rationality, concept (application) 103
 democratic rationalization 102–4, 273–4
 Denial of Service (DoS) 143
 “De Optimo Genere Oratorum” (Cicero) 219
 depth-first approach 199
 Derrida, Jacques 87–8
 descriptive work 199–200
 design 208; assumptions 213–14; case study
 211–13; C-K theory 204; contrast 214;
 engineering design 196–202; epistemology
 209–10; execution, requirement 208; failure
 210; form/function 209; research 204; science
 204–5; social aspects 210–11; thinking 196,
 202–4
Design for the Base of the Pyramid 201
Design for X 201
Designing Engineers (Bucciarelli) 210
 Dewey, John 88, 252
 Dharma-dao religions, transcendence/
 techno-lifeworld 94–5
 Diagnostic and Statistical Manual of Mental
 Disorders 178
 dialog boxes, usage 188
 Dick-Reed, Grantley 231
 digital rights management (DRM) 131, 136
 digital technoart 59–60
 digital technology: dissemination 131; usage 66
 digitization, progress 216
 discriminating taste 15
 discrimination 14; cognitive capacity 76
 discriminatory artifacts 18–25
 discriminatory technologies: defining 15–18; full
 theory, examination 14–15
 distinction, principle 76

- Distributed Denial of Service (DDoS) 143, 146, 149; attack, success 149; injection 147
 “Do Artifacts Have Politics” (Winner) 14
 domination 46–7
 Dreyfus, Hubert 45, 84–5, 89, 177, 199, 244
 Drucker, Peter 252
Dual Nature of Technical Artifacts program 198
 DuBois, W.E.B. 17, 20
 Duke University, Center on Global Change 277
 Dunbar, Robin 121
 Dutch Transition Management 259
 Dyer, Richard 14
 dysconscious racism 20
- earth, leaving (economics) 309–10
 Earth Day 101
 economic approach, critique (quantitative methods) 266
 economic-based analyses 271
 economic methods, critique 267–72
 economic models, cultural models (copyright) 130
 economy, culture (contrast) 136–8
 education 83–5; computers, usage (criticism) 85; technological approach 87–90
 educational technology 82, 89; computers/simulation/education 83–5; Pharmakon 85–7
 Eggers, Dave 42
 Electrical Numerical Integrator and Computer (ENIAC) 6
 electric amplification 59
 electronic fetal monitoring (EFM) 230
 Eliade, Mircea 92
 Emanuel, Kerry 267
 embodiment: relations 186–7; technics 19–21
 emerging technology, ethics 110
 emotional contagion study, framing 165–7
 End User License Agreements (EULAs) 158, 160
 energy: consumption, good life (relationship) 297–9; demand, estimates 268–70; International Energy Agency (IEA) research 268; society development perspective 299–300; system, sociopolitical structure 255
 Energy Paradox 263, 267, 270–2
 enframing 224–5, 228–9; phenomenology 225; reproductive enframing 230
 engineering: design 196–202; design thinking 203; thinking 203; well-ordered engineering 314
 Engineering and Physical Sciences Research Council (EPSRC), SPICE grant proposal 269
Engineering for Change (Wisnioski) 317
 engineers, collaboration 211
 entities, disclosure 225
 environment: climate models, projections 269; economic methods, critique 267–72; informational environments 124; issues, economic studies critique (assessment) 266–7; space environment 332–3
 environmental movement 253
- epidurals, usage 230–1
 epistemology: design epistemology 209–10; issues 176–7; space 341–2
 ethics: frameworks 334–5; issues 177–80; video games, relationship 29; virtue ethics 36
 etiquette, breaches 119
 eudaimonia 36
 executive system exploits 143–4
 existence, *samsara* 95
 existential risk, armed robots (relationship) 78–9
 existential safety 329–30
 Experiments in Art and Technology (E.A.T.) 68
 Expert and Citizen Assessment of Science and Technology (ECAST) network 318
 extractivism 254
 extraplanetary moral obligations/rights 330–1
 extravirtual consequences 178–9
- Facebook 42, 46, 121; defense 164; fasts 161; friends 160; Graph API 25; user grants 158–9
 face-to-face sociality 44–5
 fair information practice principles (FIPPs) 128
 fairness 169, 170–2
 Faraday, Michael 210
 Federer, Roger 185, 193
 “Federer as Religious Experience” (Wallace) 193
 Feenberg, Andrew 47, 98, 264, 273; philosophy 276–7; technology, philosophy approach 273–8
 Feyerabend, Paul 259–60
 fiction: proper fiction, utility 308; science fiction (SF) 305–8
 field composition 191–2
 field of awareness 189, 191
 field of consciousness 189
 field theory 189
 filter bubble 24
Filter Bubble, The (Pariser) 170
 fixed design problems 202
 Flash Crash (2010) 79
 Flickr 160
 Floridi, Luciano 15, 123–4
 Florman, Samuel 95
 footprint minimization 296–7
 forefrontedness 190–1
 forefronting 190–1
 Foucault, Michel 135, 161
Fountains of Paradise, The (Clarke) 307
 Fourth Amendment: mosaic theory 126; privacy protection, limits 125
 fracking 281; conceptual/historical overview 281–2; real-world experimentation 285–8; science, politics 282–5
 fragmentation, personalization (impact) 170
 Frankfurt School 102
 Franklin, Benjamin 95
 Franssen, Maarten 200
 Friedman, Batya 201
 “Friendly AI,” development 79
 Friesen, Norm 82–5, 90

- Frischmann, Brett 137
 Fritzsche, Albrecht 216
From the Earth to the Moon (Verne) 307
 full-body immersion 174–5
Full Spectrum Warrior 34
 functional chips, yield 242
 functional morality 74
 Future of Life 78–9
- Gabriels, Katleen 42
 Galison, Peter 56
 Galle, Per 198
 game-changers 203–4
 Gamer's Dilemma 37
 Gardiner, James 127
 Gardiner, Stephen 293
 Garnar, Andrew Wells 82
 Geerts, Robert-Jan 292
 Gehl, Robert W. 4
 gender 3; role, surfacing 5–7
Gender and Technology in the Making (Cockburn/Ormrod) 6
 Gene's Law 242
 Geneva Convention 76
 geoengineering 263, 264–6; demand estimates 268–70; economic approach, critique (quantitative methods) 266; economic methods, critique 267–72; environmental issues, economic studies critique (assessment) 266–7; social/technical criteria 276–7
 "Geoengineering the Climate" (Royal Society) 275
geschichtlich 224–5
Gestell (enframing) 224
 Glorious Revolution 252
 god-based transcendence 94
 Goddard Institute for Space Studies (NASA) 265
 good life 49; energy consumption, relationship 297–9
 Google 42, 46; Google Books project 6; Google Glass 121, 176, 185, 193–4
 Government Communications Headquarters (GHCQ) 162
 GPS monitoring 126
Grand Theft Auto (GTA) 30, 38
Grand Theft Auto V 30
Grand Theft Auto: Vice City 30
 Grant, Jaime M. 10
 greenhouse gas (GHG): emissions 269; production 96
 Grossman, David 34
 Group Insurance Commission (GIC), records release 167
 Grunwald, Armin 201
 Gurwitsch, Aron 189
- Halo* 2 34
 Halverson, Richard 28, 82
 Hanh, Thich Nhat 95
- Hansell, Mike 112
 Haraway, Donna 109
 Hardin, Garrett 294
 Harding, Sandra 7
 Hardt, Michael 138
 Harrison-Quintana, Jack 10
 Hartzog, Woodrow 119
 Harvard Mark I computer 209
 Hawking, Stephen 78–9
 Hayles, Katherine 65
 Hazelrigg, George 200
 HealthKit (Apple) 7
 Heidegger, Martin 16–18, 96, 154, 187, 190, 224–6; enframing concept 224, 228–9; framework 63; "Questions Concerning Technology" 224–5; "Science and Reflection" 226; technology 62–3; Western metaphysics critique 162
 Heim, Michael 177
 hermeneutic artifacts 22
Hermeneutic Phenomenology (Ihde) 55
 hermeneutic relations 187–8
 hermeneutic technics 21–3
 Hetcher, Steven 138
 Heymann, Laura 135
 Hickman, Larry 88–90, 252, 261
 Hinduism 92
 historical bias, recognition 225
 historically Black colleges and universities (HBCUs) 22
Hitting the Brakes (Johnson) 211
 Hockney, David 58
 Hoffmann, Anna Lauren 3
Holding on to Reality (Borgmann) 83
homo faber 95
homo sapiens, arrival 56
How Experiments End (Galison) 56
 Hoyt, Diana P. 326
 Hughes, Thomas Parke 211, 258
 Hull, Gordon 130
 human destiny, sharing 330
 human enterprises, animality 109
 human error, minimization 214
 humanity, art (defining feature) 63
 "humanity at work" 110
 humans, understanding 111–12
 human-technology relations 186, 190
 hydroelectric capacity (Norway) 255
 hyperreality, notion (Baudrillard) 175
- IaaS 157
 IBM, Automatic Sequence Controlled Calculator 209
Ice Age Art 55, 57
 ice age technoart 55–6
 "I Have a Dream" (King, Jr.) 136
 Ihde, Don 19, 24, 45, 54, 176, 186; background relation 188; human-technology relations forms 189; intentional fallacy 202

- Image and Logic* (Galison) 56
 immediacy, culture 45
 immersion 174–5
 immoral behaviors, cause 34
 incentives 133–6; incentives-for-creation 134–5
 indigenous religions, transcendence/
 techno-lifeworld 93
 industrialization, description 218
 industrial structures 218
 industry, users (relationship) 217–19
 information: access 85; range (narrowing),
 personalization (usage) 170
 informational environments 124
 informational privacy 123–4
 information and communication technology (ICT)
 90, 119; companies 42–3; miniaturization/
 affordability 44; usage 82, 85
 information/matter: dichotomy 65; translation 66
 information technology (IT): infrastructure,
 investment 157; transgender lives, relationship
 7–10
 infosphere, ontological friction 123–4
 infrastructure, usage 126–8
 innocence, social networking (relationship) 43–4
 inquiry, technological limits/limitations 348–9
 Instagram 160
 institutional review board (IRB) 167; scrutiny 165
 instrumental value 334–5
 intangible goods 131, 137
 intellectual property (IP) 113, 130;
 incentives-based model 136–7; proprietization
 137; protection 132–3
 interaction design 201
 Intergovernmental Panel on Climate Change
 (IPCC) 269–70; climate change discourse
 292–3; socioeconomic issue integration 266
 interlocking paths 230–1
 internal marketing research 165–6
 International Energy Agency (IEA) energy
 research 268
 international humanitarian and human rights law
 (IHL/IHRL) 73
 International Technology Roadmap for
 Semiconductors (ITRS) 240–1
 Internet Gaming Disorder 178
 Internet of Things (IoT) 145
Interstellar (SF film) 311
 interstellar settlements 335
 intravirtual consequences 178–9
 intrinsic value 334–5
Introduction to Metaphysics (Heidigger) 154
 inviolate personality 124
Ion (Plato's dialogue) 219
 iron cage 104; establishment 102
 Iron Dome (Israel) 74
 Islam 94

 Jantzen, Benjamin C. 141
 Jaspers, Karl 93

 Jenison, Tim 58
 Jenner, Caitlyn 3
 Jensen, Stine 48
 Joas, Hans 90
 Johnson, Ann 208
 Johnson, Deborah G. 127, 164
 Johnson, Jeffrey Alan 9
 Jones, Christopher 261
 Judaism 94
 Just War Theory 73, 151; armed robots,
 relationship 76–7

 Kant, Immanuel 31, 96, 120
 Kantian grounds, identification 31
 Kapp, Ernst 64
 Katyal, Sonia 134
 Katz-Rothman, Barbara 227, 231
 Keen, Andrew 45
 Keith, David 267
 Kierkegaard, Søren 84
 Kilybayev, Talgat 90
 kinetic attack 148
 kinetic warfare 149
 King, Jr. Martin Luther 136
 Kitcher, Philip 314–17
 Kitchin, Rob 4
 Klein, Naomi 254
 Kluver, Billy 68
 Knight Capital Group, stock meltdown
 (2012) 79
 know-how 227
 knowledge: mark 342–4; power, connection 98;
 propositional knowledge, explanation
 209–10
 Kranzberg, Mevlin 108
 Kroes, Peter 209
 Krueger, Myron 66
 Kurzweil, Ray 96, 238, 244–5
 Kwall, Roberta 133

 labor, disembodied experience 229
 Lamaze, Fernand 224, 231–3
 Landström, Catharina 10
 Lane, Lee 274
 Lanier, Jaron 160, 174
Laozi (opening) 95
 large-scale technical macro-systems 101–2
 Latour, Bruno 259
 legal theory, obscurity (relationship) 124–5
 Lemley, Mark 137
 lethal autonomous robots (LARs) 73; human
 supervision 75–6
 leveling down 17
 liberation 46–7
Listening and Voice (Ihde) 54
 lithotomy 228
 LNK file, shortcut 144
 loan default, prophecy 171
 Losh, Elizabeth 10, 11

- Lovins, Amory 253–5
 lows of armed conflict (LOAC) 73
 lynching game 37
- McCormick, Matthew 36
 McIntosh, Peggy 18
 McLean, Margaret 334
 McLuhan, Marshall 63–5
mala en se qualification 74
 malicious code, spread 141
Manhunt 30
Manhunt 2 30
 manipulation, concept (complexity) 166
 Mann, Jessie 62
 mappings, understanding 220–1
 market-based incentive structure 132
 market-based solutions 253–4
 Mars ClickWorkers project 132
 Mars extensibility 320
Martens Clause violation 74
 Marx, Karl 95
material prima 63
 mediated sociality 44–5
 mediation, notion 185–6
Medieval Technology and Social Change
 (White, Jr.) 58
 membrane, description 175
 mental privacy 123
 Merleau-Ponty, Maurice 190, 192
 metadiscipline 99
 meta-variable, usage 192
 Metcalfe's Law 242
 microaggressions 25–6
 micro-political activism 104
 Microsoft: Azure 158; Hololens 176;
 PowerPoint 191
 midwives, impact 227
 Mitcham, Carl 92, 218
 Model T, story 213–14
 Mody, Cyrus C.M. 238
 Mollick, Ethan 239
 Moog, usage 60
 Moore, Devin 33
 Moore, Gordon 238
 Moore's Law: contents/function 238–9; defining
 239–43; heterogeneity 242; human construct
 240–1
 Moore's Regula 238; impact 243–6
 moral consequence 109–10
 morality: functional morality 74; operational
 morality 74–5
 moral philosophy 111
 Moravec, Hans 238
 Morozov, Evgeny 45, 126–7
 mosaic theory 126
 Mount Pinatubo, eruption 265
 multiuser VR, impact 178
 Musk, Elon 78–9, 327–8
 mutable code, spectrum 144
- Nagel, Thomas 110
 nanotechnology 113
 Napoleonic Wars 219
 NASA: Asteroid Redirect Mission (ARM) 317;
 Asteroid Redirect Mission (ARM), pTA
 318–19; Goddard Institute for Space Studies
 265; participatory technology assessment 314;
 planetary defense 319–20
 National Research Council, cyberattack
 definition 148
 National Security Agency (NSA) 146, 162, 243;
 surveillance practices 43
 National Transgender Discrimination Survey
 (NTDS) 10
 natural philosophy 93
 natural science: empirical impetus 346–8;
 technology, role 346
 necessity, principle 76; requirements 77
 Needham, Joseph 95
 negative externalities 137
 Negri, Antonio 138
 Nehamas, Alexander 87
 networked communication 174–5
 networked privacy 123
New Atlantis (Bacon) 100
 New Axial Age 96
 New Caledonian crows 112
 Niaux Cave, France 55–6
 Nietzsche, Friedrich 36
1984 (Orwell) 162
 Noble, David 310
 Noble, Safiya 10
 noncommunicative interaction chains, usage 146
 non-humans, purposeful activity 110
 nontechnological childbirth 224
 nonvoluntary self-disclosure 177–8
 Norman, Donald 200–1
 Nozick, Robert 175–6
 Nussbaum, Martha 138
- objective knowledge 315
 object worlds 211
 obscurity 119; breach 122; continuum 120–1;
 fundamentals 119–22; law 125–6; legal
 theory, relationship 124–5; obscurity-corrosive
 practices 128; obscurity-related issues 120;
 online protection, code/infrastructure (usage)
 126–8; practical obscurity 127; probabilistic/
 diminishing levels 122; protections, Sotomayor
 perception 126; technology philosophy,
 relationship 123–4
 obscurity-corrosive innovations 120
101 Design Methods (Kumar) 198
 O'Neill, Onora 109
 one-world ontology 95
 online recommender system (Amazon) 4
 ontological untrustworthiness 177
 ontology: issues 175–6; one-world ontology 95
 opacity 82

- operating system (OS) 144
- operation, semi-autonomous modes 75
- operational morality 74–5
- opinions, disagreements 100
- orbital debris 320
- ORLAN 67
- Ormrod, Susan 6
- Orwell, George 162
- Ox, Jack 66

- packets, message transfer 145
- Page, Larry 6
- Painless Childbirth* (Lamaze) 231–2
- pain relief, call 226–31
- painting, instrument 63
- paleolithic technoart 55–6
- Palin, Sarah (email access) 159
- panopticon, metaphor 168
- Pariser, Eli 24, 170
- Parkin, Simon 30
- participatory democracy 101
- participatory design 201
- participatory technology assessment (pTA) 314, 317–22
- Partridge, Stephanie 37
- Peirce, C.S. 200
- perceptual gestalt, formation 187
- Perfect Moral Storm, A* (Gardiner) 293
- performance society 47
- personalization 169–70
- person of color (PoC): identification 16; selection 23; selfies perception 20
- Petrik, James 90
- Phaedrus* (Plato) 83, 86
- Pharmakon, educational technologies (relationship) 85–7
- physical practices, absence 177
- physical privacy 123
- Pinch, Trevor 60
- Pirtle, Zachary 314
- Pitt, Joseph C. 108, 110, 340
- planetary defense 319–20
- planetary protection/ethics 330–1
- planetary settlements 335
- plans, selection 87
- Platform as a Service (PaaS) 157–8
- Plato 87–8, 219; assumptions 89
- political agency 98
- political citizenship 99
- Political Reform Act of 1974 (California passage) 127
- Posey, Codey 33
- “Post Human” show 67–8
- Postman, Neil 155
- postphenomenology 185
- power: knowledge, connection 98; role 98–9
- practical obscurity 127
- praxis 54
- predictive algorithms, usage 169
- predictive policing 172
- pregnancy score 169
- Primate Visions* (Haraway) 109
- privacy 119, 167–9; decisional privacy 123; informational privacy 123–4; interests/harms 119; intimacy 122; loss, harm 167–8; mental privacy 123; networked privacy 123; physical privacy 123; prediction/fairness 169–72; privacy by design 128; protection, anonymization (usage) 167; protections, continuist interpretation 124; understanding 167–8
- private life, digitized technology (permeation) 222
- private property, perspective 263
- private publics 120
- private self 45–6
- privilege, definition 18
- problems (finding), engineering design (usage) 198–200
- production: non-market-based forms 132; treadmill 254
- Programmable Logic Cotroller (PLC) 144, 150
- Project Aurora 142
- proper fiction, utility 308
- property 133–6
- proportionality, principle 76–7; requirements 77
- propositional knowledge, explanation 209–10
- Protestant Reformation, ideological battles 155
- prototypes, behavior 212
- public interventions, interaction 106
- public self 45–6
- Pugh matrices 200
- purified essence 62
- Pursell, Carroll 108

- Quality Function Deployment 200
- quasi-otherness 188
- “Question Concerning Technology” (Heidegger) 224–5

- racial identification 16
- Rackspace 158
- radical design 198
- radio, invention 59
- radio-frequency waveforms, sensors 146
- Randall, Alice 138
- Raspberry Pi 145
- rationality: bounded rationality 123; contesting, analysis 200; democratic rationality 103; pure rationality 104–5; technical rationality 104
- rationalization 102, 104; democratic rationalization 102–4
- Raustiala, Kal 132–3
- Read method, basis 232
- reality: feedback 106; fullness 177
- real-world environment, mapping (establishment) 218

- recording devices, invention 59
Recreating Motherhood (Katz-Rothman) 227, 231
 reframing design, phenomenon 199
 Regan, Priscilla 127
 relations: alterity relations 23–4; background relations 24–5
 religion: Abrahamic religions, transcendence/techno-lifeworld 93–4; dharma-dao religions, transcendence/techno-lifeworld 94–5; indigenous religions, transcendence/techno-lifeworld 93
Religion of Technology (Noble) 310
 religious transcendence 92
 Renaissance Perspective 58
 Renaissance technoart 58–9
 renewable energy: distribution 256; green politics, association 253; intermittent usage 255; myths 254–7; progressive integration 260; scaling 256–7; supply 255; technology, principles 256–7; technology, revolution 251; visibility 257
 repair mechanisms 244
 reproductive enframing 230
 Rescher, Nicholas 346
research through design 204
 resource attacks 143
 “Return to Innocence” (Enigma) 135
 revolution, meaning 252
 revolutionary technology, rethinking 258–61
 Rheingold, Howard 177
 Rich, Adrienne 228
 Riskin, Jessica 113
 Rivera, Ignacio G. 10
 Robert Bosch GmbH 213
 robotic actions, rubber-stamp approvals/post-hoc justifications 75–6
 robotics, three laws (Asimov) 78
 robots: lethal autonomous robots (LARs) 73; militarization 73; philosophical questions 73
 Rock’s Law 242
 Rosenberger, Robert 90, 185
 Rothenberg, David 64
 Rowlands, Mark 64
 RQ-170 drone, downing 148
 Rushkoff, Douglas 161

 SaaS 158
 Sacco, Justine 46
 sacred-secular culture, integration 93
samsara 95
 Sandy Hook Murders 33
 Sartre, Jean-Paul 96
 satisficing 197
 Schleiermacher’s text, interpretation 219
 Schnaiberg, Allan 254
 Schön, Donald 199
 Schulzke, Marcus 31–2, 35, 38
 Schumacher, E.F. 256

 science: alien science, potential 349–51; design/production, gender role (surfacing) 5–7; policy, objective knowledge/democratic input 315; social science, well-ordered science (connection) 316–17; well-ordered science 316–17
 “Science and Reflection” (Heidegger) 226
 science fiction (SF) 305; earth, leaving (economics) 309–10; social/technological change, reaction 306–8; space tourism, combination 311–12
Science & the Age of Sensibility (Riskin) 113
Secret Knowledge (Hockney) 58
 secular transcendence 95–6
 sedimentation 192–3
 self, contrasts 45–6
 self-destruct code 128
 self-neglect, moral problem 32
 Self-Strengthening Movement 95
 Selinger, Evan 119
 Sellars, Wilfrid 113
 semiocapitalism 47
 Sen, Amartya 138
 Serial Peripheral Interconnect (SPI) 145
 service, denial 143
Shamans, Software and Spleens (Boyle) 135
 sharing: algorithmic culture 47–9; dialectics 42
 sharing economy 26
 Shew, Ashley 108
 Shormanbayeva, Dinara 90
 Sikka, Tina 263
 Silbey, Jessica 133
 Simon, Herbert 196, 244
 simulation 83–5
 Singer, Peter 109
 Singularity 244–5; theories 78
 Singularity Universe 245
 Siri, usage 244
 skidding, phenomenon 212–13
 Skynet scenario 78
 Smellie, William 227
 Smith, Cyril Stanley 62, 66
 Snowden, Edward 49
 social actors 105
 social brain hypothesis 121
 social change, reaction 306–8
 sociality 44–5
 social media vacations 161
 social networking 42; innocence, relationship 43–4
 social networking sites (SNSs) 42; description 43; liberation/domination 46–7; system, analysis 46–7
 social relations 134–5
 social science, well-ordered science (connection) 316–17
 social surveillance 123

- society: command society 47; conquest 101; development, energy perspective 299–300; performance society 47
- socioepistemological activity 211
- Socrates 86
- software problem 239
- solar electric propulsion (SEP) 319
- solar radiation management (SRM) 264, 272, 274
- solutions: finding, engineering design (usage) 197–8; generation 203–4; market-based solutions 253–4; principle 197
- Soraker, Johnny Hartz 174
- Sotomayor, Sonia 126
- soul, selection 87
- Souls of Black Folks, The* (Du Bois) 17
- space: debris 333; environment 332–3; epistemology 341–2; preservation 334; property rights/resources 331–2; stewardship 334; transcendence 340
- space exploration 329–30; cognitive prosthesis 337–8; commercialization/exploitation 336; cyborgs 337–8; ethical frameworks 334–5; humachines 337–8; human purpose 330; justice/governance 335–6; planetary/interstellar settlements 335; risks 335; shared human destiny 330; stakeholder interests 336–7; survival/existential safety 329–30; value propositions, negotiation 336–7; values 334; worth 334–5
- space tourism 305; earth, leaving (economics) 309–10; motivations 310–11; science fiction, combination 311–12
- Spade, Dean 9, 11
- Sparrow, Robert 332–3
- SPICE *see* Stratospheric Particle Injection for Climate Engineering
- spirit/matter, dichotomy 65
- Sprigman, Christopher 132–3
- stakeholder interests 336–7
- Stallman, Richard 161
- standing-reserve, objectlessness 229
- Star Trek: Reboot* 135
- Steadman, Philip 48
- Strahilevitz, Lior 132
- Stratospheric Particle Injection for Climate Engineering (SPICE): Deliberative Workshops 270; project 263–6, 274
- Strickland, Sunnys 229
- structural rights 124–5
- Structured Query Language (SQL) 143; injection 147
- structure/function, dual nature 209
- Stuxnet worm 142, 144
- subject-object distinction 233
- Sullivan, Louis 208
- Sunder Madhavi 134
- Surden, Harry 124–5
- survival, safety 329–30
- Swales, Sinqe 34
- Sweeney, Latanya 167
- Swierczek, Julie C. 154
- tablet screen, buttons (appearance) 220
- Tacoma Narrows bridge, failure 210
- Tallinn Manual 148
- Target, algorithms (usage) 169
- Techné* (Hickman) 252
- techne 227; technology, connection 87
- technical citizenship 98–100
- technical code 105
- technical disciplines: interaction 106; responses 105–6
- technical feasibility 217–18
- technical rationality 104–5
- technics 93; embodiment technics 19–21; hermeneutic technics 21–3
- technoart 54; African technoart 56–7; beginnings 56; digital technoart 59–60; ice age technoart 55–6; image/logic 56; paleolithic technoart 55–6, 57; Renaissance technoart 58–9
- technocratic ambition 100–2
- technocratic ideology 98
- technodelics 95
- techno-lifeworld 93–5
- technological achievement 245
- technological childbirth 224
- technological fix 203–4
- technological information 83–4
- technologically mediated experience: alterity relations 188; background relations 188; cyborg relations 189; embodiment relations 186–7; field composition 191–2; forefronting 190–1; form 186–9; hermeneutic relations 187–8; organization 189–93; sedimentation 192–3; transparency 190
- technological model, technocratic characteristic 224
- technological revolution 261
- technological society, agency/citizenship 98
- technological unconscious 20
- technology 3; American educator criticism 82–3; animal concept, absence 112–13; anthropocene usage 300–1; artists, relationship 62; bidirectional flow 65; Big Technology 113; category 111, 112; change, reaction 306–8; common-sense definition 23; contemporary body 67; critical theory 104–6; design/production, gender role (surfacing) 5–7; discriminatory technology 14–18; educational technology 82; engineer design 222; free relationship 231; information technology, transgender lives (relationship) 7–10; instrumental understanding 226; paraphrases 221; participatory technology assessment 314; philosophers, reference 110; real world (Mitcham) 218; revolutionary technology,

- rethinking 258–61; role 346; shaping, artists (impact) 68; techne, connection 87; technology-informed approaches, philosophy 296–300; theory (Heidegger) 224–6; Theuth, relationship 86; trajectory, transcendence 96–7; translational approach 216; translations 221; valuative bias 105
 technology, philosophy 62, 83, 259; animals 108; approach (Feenberg) 273–8; artifacts 300; climate change, relationship 292, 300–1; Moore's Regula, impact 243–6; obscurity, relationship 123–4
Technology and the Lifeworld (Ihde) 176
Technology Assessment (Grunwald) 201
 technology development 319; dynamics, change 216–17
 technophilic childbirth 226–31
 technophobic childbirth 231–3
Technopoly (Postman) 155
 telepresence 174–5
 tennis, wearable technology 193–4
 Terms of Service (ToS) 47
 terraforming 332–3
 Theremin, Leon 59–60
 Theuth, technology (relationship) 86
 third party doctrine 125
 "This Is Water" (Wallace) 193
 Thomson, Iain 225
 thought, artist translation 66
 Three Laws of Robotics (Asimov) 78
 Tomblin, David C. 314
Topics in Cognitive Science 64
 transcendence 92–5, 340–1; god-based transcendence 94; secular transcendence 95–6; speculative conclusion 96–7
 transgender, defining
 transgender lives: consideration 3; data/information technology, relationship 7–10
Transgender Studies Quarterly 9–10
 translation 66, 219–20; cultural approach 221–2; distance, establishment 219
 Transmission Control Protocol/Internet Protocol (TCP/IP) 145
 transparency 47–9, 190; social interactions 48; type 42
 Trocco, Frank 60
 Turing Test, passage 243–4
 Turkle, Sherry 45, 177, 179–80
 Tushnet, Rebecca 133
 Tvindmill 251
 Tvind School 251
 twilight sleep, usage 229
 Twitter Revolution 47
 Twombly, Cy 63
 Uber service 26
 Uniform Record Locator (URL) system, access 145
 United Kingdom Energy Research Centre (UKERC) 271
 United Nations Framework Convention on Climate Change (UNFCCC) criteria 277
United States v. Jones 126
 unmanned aerial vehicle (UAV) 74
 Unmanned Aerial Vehicle (UAV), signals 147
Uplife Universe (Brin) 306
 Upturn Report 171
 usability, focus 201
U.S. Department of Justice v. Reporters Committee for Freedom of the Press 125
 user-centered design 201
 users: designs/assumptions 213–14; experience, organization 185; field of awareness 189; industry, relationship 217–19
 Vallor, Shannon 73, 178
 value propositions, negotiation 336–7
 values: designing, possibility 202; disagreement 100; realization, engineering design (usage) 200–2
 Van Den Eede, Yoni 42, 48
 Verbeek, Peter-Paul 186, 189
 Vermaas, Pieter E. 196
Vermeer's Camera (Steadman) 58
 Verne, Jules 307
 video games: ethics, relationship 29; moral controversy 29–30
 Vincenti, Walter 258–9
 Viola, Bill 64
 violent video games: behavioral effects 31–4; characters, relationship 36–7; concerns, addressing 37–8; moral significance, assessment (approach) 31–7; playing, moral fishiness 34–5; psychological effects 34–6
 Virgin Galactic, space line 311–12
 Virginia Tech Massacre 33
 virtual ATMs 176
 virtual communities 43
 virtual environments 174; epistemological issues 176–7; ethical issues 177–80; ontological issues 175–6; reality 175
 virtuality fallacy 175
 virtual lighting 176
 virtual reality (VR): exposure therapy 179; metaphysics 174–5; multiuser VR, impact 178; research 174; usage 179
 virtue ethics 36
 Waddington, David 31–2, 35
 Wagner, R. Polk 137
 Wallace, David Foster 186, 193
 War on Christmas 22
 Washington Naval Yard Shootings 33
 Watson (IBM) 244
 Watson, Matt 265
 Wayland, Kent 127

- wearable technology 193–4
- Web 2.0, rise 131–2
- Weber, Max 102
- webpages, trackers 24–5
- Weld, William 167
- well-ordered engineering 314
- well-ordered science (WOS) 316–17, 316t
- We Sing the Body Electric* (Anstey) 67
- Westin, Alan 121–2
- West, Kanye 135
- what-it's-like 21
- White, Jr. Lynn 58, 94
- White Is* (Wilcox) 14
- whiteness, global assumption 20
- Wilcox, Preston 14
- Williamson, Mark 334
- wind development, traditional model 256
- Wind Done Gone, The* (Randall) 138
- wind turbines: design/performance, technical questions 258; impact 258
- Winner, Langdon 14, 19, 98, 253, 258
- Wisnioski, Matthew 317
- Wittkower, D.E. 14
- Wonderly, Monique 29, 35
- Woodruff, Paul 87
- world, resource 224–6
- writing, technological aspect 87
- Wylie, Alison 7
- Yir Yoront peoples, anthropological story 93
- You Are Not a Gadget* (Lanier) 160
- Young, Garry 35
- Zehou, Li 95
- zero-day exploits 144
- Zhai, Philip 177