VALUE SENSITIVE DESIGN

SHAPING TECHNOLOGY WITH MORAL IMAGINATION

BATYA FRIEDMAN
DAVID G. HENDRY

Value Sensitive Design

Shaping Technology with Moral Imagination

Batya Friedman and David G. Hendry

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in ITC Stone Sans Std and ITC Stone Serif Std by Toppan Best-set Premedia Limited. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Friedman, Batya, 1957- author. | Hendry, David, 1962- author.
Title: Value sensitive design : shaping technology with moral imagination / Batya Friedman and David G. Hendry.
Description: Cambridge, MA : MIT Press, [2019] | Includes bibliographical references and index.
Identifiers: LCCN 2018028524 | ISBN 9780262039536 (hardcover : alk. paper)
Subjects: LCSH: System design--Methodology. | Human-computer interaction. | Computer programming--Moral and ethical aspects. | Computer programmers--Professional ethics.
Classification: LCC QA76.9.S88 F75 2019 | DDC 004.01/9--dc23 LC record available at https://lccn.loc.gov/2018028524

10 9 8 7 6 5 4 3 2 1

Contents

Preface and Acknowledgments xv

1 Introduction 1
2 Theory 19
3 Method 59
4 Applications 105
5 Conclusion 167

Envoi I: Visual Language 181
Envoi II: Photo Poem 185
References 189
Index 219

Preface and Acknowledgments

Value sensitive design emerged in the 1990s. The roots of the value sensitive design tripartite methodology can be found as early as 1992 in a paper by Friedman and Kahn (1992) in which they engaged in a conceptual analysis of human agency and responsible computer system design. That paper provided careful analyses of agency, human and machine; implications of that agency for moral action; and considerations for technical features in computer system design. The paper concluded with a call to bring an empirical understanding of people's moral psychology into the mix. The term "value sensitive design" first appeared in the mid-1990s with a perspective in the ACM *Interactions* entitled "Value-Sensitive Design" (Friedman, 1996), followed shortly thereafter by an edited volume published by Cambridge University Press and the Center for the Study of Language and Information, *Human Values and the Design of Computer Technology* (Friedman, 1997).

Two workshops funded by the US National Science Foundation helped to cultivate a research and design community in value sensitive design. The first workshop, held in Washington, DC, in May 1999 and attended by Edward Felten, Batya Friedman (organizer), Jonathan Grudin, Helen Nissenbaum, and Terry Winograd, sought to set a research agenda for value sensitive design (Friedman, Felten, Grudin, Nissenbaum, & Winograd, 1999). The second workshop sought to further develop and refine that research agenda and to stimulate research in this approach by cultivating a broad community of value sensitive design researchers. Held at the University of Washington in Seattle, Washington, in September 2000, the second workshop was organized by Alan Borning and Batya Friedman and attended

by nearly 30 participants,[1] who, as a group, represented perspectives from cognitive science, computer ethics, computer security, computer-supported cooperative work, design, human-computer interaction, interaction design, participatory design, philosophy, social-psychological aspects of information systems, software development, technology, and ubiquitous computing. Many of the individuals who attended the 1999 and 2000 workshops have either contributed directly to the literature on value sensitive design or in other ways engaged explicitly with human values in their technical research and design work.

Also in 1999, Batya Friedman and Peter Kahn joined Alan Borning at the University of Washington, where they established the interdisciplinary Value Sensitive Design Lab. Their co-location and support from their respective units—the Information School, Department of Psychology, and Department of Computer Science and Engineering—provided a fruitful environment for developing value sensitive design.

Four prior synthetic accounts of value sensitive design provide a window onto its development. The first was a handbook chapter on human values, ethics, and design (Friedman & Kahn, 2003; rev. ed., 2008), which placed an early description of value sensitive design in a broader discussion of how values become implicated in technological design; other approaches to human values, ethics, and design; human values of ethical import; and professional ethics. The second was an entry on value sensitive design in an encyclopedia of human-computer interaction (Friedman, 2004). The third, a chapter devoted to value sensitive design that explicated some theory, three case studies, and heuristics on method (Friedman, Kahn, & Borning, 2006a; reprinted in Himma & Tavani, 2008 and in Doorn, Schuurbiers, van de Poel, & Gorman, 2013). The fourth, a survey article on value sensitive design methods that summarized core theoretical commitments, reported on 14 methods and provided heuristics for skillful value sensitive design practice (Friedman, Hendry, & Borning,

1. The following individuals participated in the 2000 workshop: Gregory Abowd, Alan Borning (organizer), Tone Bratteteig, Philip Brey, Gary Chapman, Edward Felten, Raya Fidel, Batya Friedman (organizer), Jonathan Grudin, Jim Gray, Chris Hoadley, Peter Kahn, Wendy Kellogg, Jennifer Mankoff, Elizabeth Mynatt, Clifford Nass, Helen Nissenbaum, Andreas Paepcke, Kurt Partridge, Karen Pettigrew, Steven Poltrock, Lodis Rhodes, Ole Smørdal, Deborah Tatar, John Thomas, Peter Paul Verbeek, Paul Waddell, and Terry Winograd.

2017). That survey article has been expanded and is now chapter 3 in this book.

Moreover, during the seven-year period of writing this book, review articles have also appeared in the literature, including one by Davis and Nathan (2014), who discuss applications, adaptations, and critiques of value sensitive design, and one by Huldtgren (2015), who reviews some of the methods and theoretical aspects of value sensitive design. Having received doctoral or internship training at the Value Sensitive Design Lab at the University of Washington, Davis, Nathan, and Huldtgren offer distinctive and insightful points of view in their review articles.

Most recently, two new workshops (Friedman et al., 2015; Friedman, Harbers, Hendry, van den Hoven, & Jonker, 2016) have taken up the charge of framing grand challenges for value sensitive design going forward. The first, titled "Charting the Next Decade for Value Sensitive Design," was a one-day workshop held in August 2015 at the Fifth Decennial Conference on Critical Alternatives in Aarhus, Denmark. Organized by Batya Friedman, David Hendry, Jeroen van den Hoven, Alina Huldtgren, Catholijn Jonker, and Aimee van Wynsberghe, the workshop brought together 19 researchers and designers from such fields as computer science, engineering, human-computer interaction, law, library and information science, and philosophy.[2] Focusing on human values and technology, workshop participants began a conversation on grand challenges for value sensitive design. A year later, in November 2016, the Aarhus conversation continued at a five-day workshop at the Lorentz Center in Leiden, The Netherlands. Similarly titled "Value Sensitive Design: Charting the Next Decade," the Leiden workshop was organized by Batya Friedman, Maaike Harbers, David Hendry, Jeroen van den Hoven, and Catholijn Jonker.[3] There were 41 participants in

2. The following individuals participated in the 2015 Aarhus workshop: Anette Andersson, Batya Friedman (organizer), Maaike Harbers, David Hendry (organizer), Jeroen van den Hoven (organizer), Alina Huldtgren (organizer), Sampsa Hyysalo, Catholijn Jonker (organizer), Michael Katell, Alex Kayal, Ian King, Lisa Nathan, Bryce Newell, Rose Paquet Kinsley, Jeremy Pitt, Luke Stark, Åke Walldius, Daisy Yoo, and Bieke Zaman.

3. The official website for the Lorentz Workshop, "Value Sensitive Design: Charting the Next Decade," can be found here: https://www.lorentzcenter.nl/lc/web/2016/852/info.php3?wsid=852&venue=Oort.

Leiden,[4] many of whom had also attended the Aarhus workshop. Along with paper presentations and numerous hands-on activities, distinguished conversations were held with Lisa Nathan, Sarah Spiekermann, Alan Borning, and Volker Wulf. The 2017 Lorentz workshop resulted in new research nuggets on recent work in value sensitive design and a set of grand challenges for moving the field forward over the next decade. These workshop outcomes will be published in a special issue of the *Journal of Ethics and Information Technology* (forthcoming).

Turning now to specific acknowledgments: many, many people have contributed to bringing value sensitive design to its current form. To all those, we are enormously grateful.

Early supporters who challenged, critiqued, and enabled the work include Ron Baecker, Suzanne Iacono, Sara Kiesler, Rob Kling, Ben Shneiderman, and Terry Winograd. Early collaborators include Sunny Consolvo, Edward Felten, Ken Goldberg, Peter Kahn, Clifford Nass, Helen Nissenbaum, Ian Smith, and John Thomas.

Alan Borning: no other person has contributed so much to value sensitive design over the years—intellectually, practically, and always with integrity. To Alan, first and foremost dear friend: deep respect and appreciation.

Our former students, now friends and colleagues—Norah Abokhodair, Janet Davis, Tamara Denning, Katie Derthick, Abigail Evans, Nathan Freier, Daniel Howe, Shaun Kane, Travis Kirplean, Predrag Klasnja, Milli Lake, Peyina Lin, Jessica Miller, Lynette Millett, Lisa Nathan, Bryce Newell, Trond Nilsen, Kari Watkins, Jill Woelfer, and Daisy Yoo—have contributed substantially to the development of value sensitive design while pursuing their own research interests and successful careers. Some of their work appears

4. The following individuals participated in the 2016 Lorentz Center Workshop: Tag Alshehri, Christian Bonnici, Alan Borning, Oliver Burmeister, Christian Detweiler, Batya Friedman (organizer), Christiane Grünloh, Maaike Harbers (organizer), Oliver Heger, Donal Heidenblad, David Hendry (organizer), Alina Huldtgren, Naomi Jacobs, Nassim Jafarinaimi, Catholijn Jonker (organizer), Ian King, Marjolein Lanzing, Qinyu Li, Nick Logler, René Mahieu, Noëmi Manders-Huits, Jason Millar, David Miller, Lisa Nathan, Bryce Newell, Bjoern Niehaves, Anne Nigten, Sarah Spiekermann, Luke Stark, Tjerk Timan, Ibo Van de Poel, Jeroen Van de Hoven (organizer), Peter Van Waart, Aimee Van Wynsberghe, Pieter Vermaas, Åke Walldius, Kari Watkins, Till Winkler, Volker Wulf, Daisy Yoo, and Annuska Zolyomi.

past and future, which in time we hope will contribute to our common humanity.

Batya Friedman and David Hendry
September 2018
The Plank Table
Zoka Coffee, University Village
Seattle, Washington
United States

1 Introduction

The water vessel, taken as a vessel only, raises the question, 'Why does it exist at all?' Through its fitness of construction, it offers the apology for its existence. But where it is a work of beauty it has no question to answer; it has nothing to do, but to be.
—Rabindranath Tagore (1922, p. 5)

In these few words, Rabindranath Tagore gently points us to the human condition. We learn from Tagore that being with our tools gives consideration not only to functionality but also to human flourishing. Thus, retelling Tagore's story of the water vessel, we might say: Every human being is entitled to clean water to drink and a vessel from which to drink that water. And that vessel should be beautiful.

Technology is the result of human imagination—of human beings envisioning alternatives to the status quo and acting upon the environment with the materials at hand to change the conditions of human and nonhuman life. As a result of this human activity, all technologies to some degree reflect, and reciprocally affect, human values. It is because of this deep-seated relationship that ignoring values in the design process is not a responsible option. At the same time, actively engaging with values in the design process offers creative opportunities for technical innovation as well as for improving the human condition.

Our human imaginations have the potential to be moral—to imagine what constitutes lives of quality and societies of quality, human beings living well and other living creatures living well. Technology shapes our human experience and impacts all of nature. Thus, in Terry Winograd and Fernando Flores's (1986, p. xi) words, "in designing tools we are designing ways of being"—ways of being with moral and ethical import.

This book is about bringing our moral and technical imaginations into the designing and making of technology writ large. It is about expanding our imaginations and opportunities, our toolsets and methods, and our criteria for judging the quality of systems we build. While empirical study and critique of existing systems is essential, this book is distinctive for its design stance—envisioning, designing, and implementing technology in moral and ethical ways that enhance our futures.

All human beings have enduring desires—to survive, to explore the natural world, to be in community, to create and experience beauty through music and art, to dream, to live in health, to prepare for death. We look inward and we look outward, sometimes narrowly and sometimes expansively. These and other desires have led to a myriad of inventions, and those inventions then shape our lives going forward. This is familiar news.

Consider these technologies. The Internet has enabled new and particular forms of remote communication and, in so doing, transformed our ideas about and how we experience friendship, parenting, caring for our elderly, community, and social networks. Sophisticated search algorithms can also bring us closer to people who appear to be and think like ourselves, wherever they are physically located. On the one hand, such connection can foster community in the face of isolation; on the other hand, it can lead to social and information "bubbles," risking isolation from people with ideas, belief systems, and ways of being that differ from our own. The Internet of Things and related technologies enable smart cities that manage electricity consumption in more energy-efficient ways, but also further dependence of cities and societies on a 24/7 power grid that, in turn, requires an energy infrastructure. Algorithmic decision-making systems press on education, work, and law as they increasingly make college entrance, hiring, firing, and criminal sentencing, parole, and other institutional decisions. Over time, biases that become embedded in such systematic decision making have the potential to magnify existing structural inequalities. Computing and information technologies wrapped in the ephemeral language of "the cloud" and "Ethernet," working at the speed of nanoseconds, give the illusion of minimal physical, material, and energy impact. Yet, in reality, these technologies possess a large footprint in the form of server farms, e-waste, and vast amounts of energy consumption to produce hardware and run complex algorithms. Indeed, the impact of human activity amplified by our technology use on atmospheric, geologic, hydrologic, biospheric, and

other earth systems has been so transformative as to lead some scientists to propose a new geologic epoch, that of the Anthropocene.

Similar observations extend to simpler and nondigital technologies used in urban or rural settings. Consider food: powdered energy drinks and other packaged foods now populate our markets, are regular parts of human diets, and have changed ideas about what it means to eat and drink healthy food. Or consider the water mill and its continued use by farmers in developing countries. Or consider durable goods and infrastructures in energy, transportation, warfare, health care, food systems, housing, water, and sanitation—such as wind turbines, bicycles, bombs, pill bottles, organic foods, energy-efficient dwellings, faucets, and plumbing. It is noteworthy that these tools and technologies may or may not be tied to information systems, as well, for example via embedded digital technology or by being part of global information-intensive supply chains. This, however, is not the main point. What matters is that a tool or technology is in an *interactive* relation to human beings. As Winston Churchill said in 1943, "we shape our buildings and afterwards our buildings shape us." We think the same is true, not just for buildings, but for any tool or technology.

There is a tightly coupled interaction between our experience of ourselves as human beings and our tools and technologies—so much so, that it makes little sense to speak of one without the other. Given this coupling, how then can researchers, designers, engineers, and policy makers engage in the design and development of tools and technology to support human flourishing in all its richness—to enable human beings to grow and develop, to make manifest what they value, and to act meaningfully and ethically in the world?

While there are surely many ways to approach a design and engineering process addressing this question and these aims, in this book we take up one approach, that of value sensitive design.

Value Sensitive Design in Essence

Value sensitive design seeks to guide the shape of being with technology. It positions researchers, designers, engineers, policy makers, and anyone working at the intersection of technology and society to make insightful investigations into technological innovation in ways that foreground the well-being of human beings and the natural world. Specifically, it provides

by the acclaimed architects Reima and Raili Pietilä and located at Aalto University's Otaniemi campus in Espoo, Finland. This was no straightforward matter, in part because the roof's silhouette, a key element of the building, might be disturbed with the installation of the solar arrays. Beginning with a conceptual investigation, the authors identified a web of values that were used to shape several empirical and technical investigations. These values included cultural heritage preservation, campus prestige and image, and ecological modernization, along with economic costs. Building on this conceptual investigation, architectural reviews and expert interviews with architects (empirical investigations) led to criteria for guiding a technical investigation. These criteria included "preservation yet modernization," "the identity of Aalto University as eco-conscious," and "producing renewable energy without greenwashing." Next, a technical investigation was conducted, where prototypes of solar arrays were placed on the roof of the Dipoli. These prototypes were studied in terms of their aesthetics and their expected energy production—values that are entwined, since the orientation of the solar arrays to the sun determines expected energy production but also impacts the aesthetics of the roof's silhouette. In summary, the integrative and iterative use of conceptual, empirical, and technical investigations led to the key concept of "subtle visibility," which refers to striking a balance between preserving the "roofscape" of the Dipoli while also allowing for substantial energy production.

These two examples are illustrative. Chapter 4 includes 10 in-depth reports of application domains, further showing how the theoretical commitments of value sensitive design can be put to practical use to meaningfully shape design processes. They show value sensitive design to be robust and actionable in its current form, while also being open to further discovery, enhancement, and improvement. The applications, as a whole, demonstrate a commitment to "progress, not perfection."

Two Hard Problems and a Way Forward for Value Sensitive Design

Value sensitive design is concerned with what people consider important in their lives, with a focus on ethics and morality. Hence, moral and ethical theory matters. Yet moral discourse is rife with disagreement. Longstanding debates about human values and ethics permeate moral philosophy

and the social sciences. In his book *Justice: What's the Right Thing to Do?*, the leading political and ethical philosopher Michael Sandel (2009) talks about the state of these debates in morality and law as revolving "around three ideas: maximizing welfare, respecting freedom, and promoting virtue. Each of these ideas points to a different way of thinking about justice" (p. 6). In the language of moral philosophy, these concerns reflect consequentialist, deontic, and virtue ethics perspectives, respectively. Adding to these Western-centric perspectives on justice, we can point to other non-Western worldviews that foreground, for example, harmony or community, as ways to organize and sanction societies. Depending on which perspective is taken, different implications for law and policy (and technology use) follow.

The complexity goes beyond abstract moral theory, connecting to the particulars of societies and technologies. To see this interconnection, think for a moment about practical systems of government and consider democracy as a specific example: the shape of communication technology that supports one form of democracy (e.g., direct democracy in a small town or city) may be quite different from that which supports another form of democracy at another scale (e.g., representative democracy spanning a continent). Moreover, even while democracies are developing, along with corresponding understandings of justice, technology is evolving, offering new ways of acting in the world.

Thus, the first hard problem: these complexities tied to moral and ethical theory are unlikely to be resolved any time soon. Technology and society will not stand still. Value sensitive design needs to chart a path forward, even while moral philosophers, legal scholars, social scientists, and others are still at work.

How then, will value sensitive design account for moral and ethical theory? It is not the intent for value sensitive design to solve these disagreements. Rather, the aim for value sensitive design is to position researchers, designers, engineers, and others to make progress in the design of technology through the foregrounding of human values even as these debates unfold and disagreements go unresolved. To do so, value sensitive design resists overarching normative directives, for example, about which ethical theory to adhere to or what particular design sequence to prescribe. That said, mechanisms exist through which normative elements can be embedded in value sensitive design. For example, a particular ethical theory could

be employed in a given project, as van Wynsberghe (2013) did in her application of care ethics to robotic assistants for health care. Or, prescriptions for carrying out specific methods could be inserted into standard engineering processes, as Spiekermann (2015) did when integrating elements of value sensitive design into a waterfall model for software engineering. Interestingly, some moral philosophers, such as van den Hoven, Lokhorst, and van de Poel (2012), suggest that design approaches like value sensitive design, which cultivate both moral and technical imagination in the context of practical activity, could contribute to the moral discourse, perhaps paving the way for new insights in these long-standing debates.

The second hard problem concerns "robustness" in terms of the scope and transferability of value sensitive design. Two questions have motivated key strategies for developing value sensitive design: "How to design an approach that can account for a wide range of values, stakeholders, technologies, populations, contexts, and circumstances, and at different scales of human experience?" and "How to develop theory and methods that can be broadly applicable, readily transferring from one situation to another?"

These considerations of "robustness" pointed us toward a pragmatic way forward, one structuring the choice and framing of value sensitive design projects as a whole. Specifically, and counter to common practice, around the year 2000 we made the strategic decision *not* to specialize in a particular value, technology, population, or context. Thus, for example, we explicitly decided not to become specialists in privacy and security, or to specialize in human-robot interaction, or to work with young children or with the elderly, or in hospital settings. Rather, our intuition was that by working across a diversity of values, technologies, populations, contexts, and levels of human experience, we would both surface blind spots in the approach (and be positioned to mitigate them) and more readily characterize theoretical commitments and methods at a level of abstraction that would be widely useful (see figure 1.1).

Following this strategy led us to explore a broad set of values and to appreciate the interconnectedness of values in human experience. It led us to adapt and invent methods that work well not only in the hands of designers but also with clients and other direct and indirect stakeholders. It led us to explore those same methods successfully with personal technologies like implantable medical devices (Denning et al., 2010) and mobile

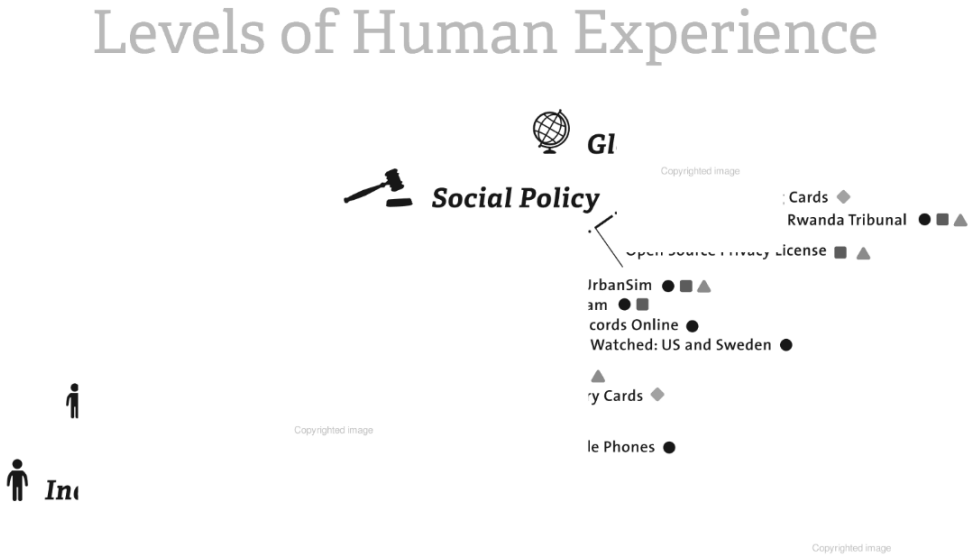

Figure 1.1
Projects that engage different levels of human experience. Slide reproduced from University Faculty Lecture, University of Washington (Friedman, 2013).

phone parenting technologies (Czeskis et al., 2010), as well as with new tools for data scientists working with machine-learning algorithms (Bender & Friedman, 2018) and information systems to support international justice (Yoo et al., 2016). Of course, the use of theory and method needs to unfold in a sensitive and responsive manner—each design situation is unique.

Value sensitive design, itself, was designed intentionally with this scope and transferability in mind: an approach that wouldn't break when presented with a new set of values, a new technology, a new population, or a new context or circumstance of use.

Related to robustness, in putting forth a new design approach with new processes, methods, and theoretical commitments, we considered carefully how best to "test" the approach and make the case for its effectiveness. Reflecting both on supporters and skeptics, we settled on the strategy of "proof-of-concept" projects that would first show, then tell, about the approach. That is, we argued for the viability of a value sensitive design

approach through the making of artifacts that employed the approach to realized desired aims. Thus, those who would challenge the approach would need to argue not only with representations on paper but also with the existence of the built artifacts themselves. This strategy proved especially effective with those who initially thought such value sensitive design couldn't be done, as the artifacts clearly demonstrated otherwise. Moreover, because we often found ourselves treading new ground—the possibility for informed consent with web browsers in early 2000 (Friedman et al., 2000a; Millett et al., 2001; Friedman et al., 2002a; Friedman, Hurley, Howe, Felton, & Nissenbaum, 2002b); the possibility for social and moral engagement with robots in the mid-2000s (Friedman, Kahn, & Hagman, 2003; Kahn, Freier, Friedman, Severson, & Feldman, 2004; Kahn, Friedman, Alexander, Freier, & Collett, 2005); the possibility for gender differences in the experience of privacy in public (Friedman et al., 2006a; Friedman et al., 2008a); and so forth—we often found ourselves at the forefront of key issues concerning the moral and ethical impact of emerging information and computing technologies.

Stepping back, scholars engaged over several decades in developing any robust body of work will likely make adjustments along the way. How could that not be the case? If it were otherwise, no new knowledge, experiences, or approaches would have been discovered, invented, or expanded. Value sensitive design is no exception. From early on, we have been alert in framing value sensitive design to articulating the approach so that, as appropriate, it could accommodate elaboration, growth, development, expansion, and discovery. Our primary strategy is to be precise about our intellectual commitments without being overly prescriptive or constraining. While much of the initial framing of value sensitive design remains intact and methods developed early on are still useful, other changes have occurred as follows: theoretical constructs have been elaborated or expanded to address gaps; some theoretical constructs have been clarified or reframed in response to new knowledge or critique; many new methods have been developed or adapted; new ways of employing earlier methods have been explored; and open questions have been articulated. As appropriate in this book, we call attention to instances for which later articulation of value sensitive design theory, method, and practice diverge from earlier accounts.

prescribe norms of behavior, as well as to explore how such innovations might extend the boundaries of traditional ethical concepts, such as privacy and agency.

Complementing this more philosophical discourse, the social informatics community (Attewell, 1987; Iacono & Kling, 1987; King, 1983; Kling, 1980), at the time not yet coalesced under this label and comprised primarily of social scientists, emphasized the sociotechnical analyses of deployed technologies. As a community, they engaged in an interdisciplinary study of the design, uses, and consequences of information technologies that took into account interaction with institutional and cultural contexts. As a group, this body of work demonstrated how the introduction of computing technology into organizations changed the nature of work, communication, commerce, education, and so forth—and, reciprocally, how individuals and organizations worked around computing/technical features to (re)assert their values. Taken together, these studies laid important groundwork for an interactional understanding of technology and human values.

However, neither the philosophical work nor the social informatics work provided much guidance for designers and engineers engaged in addressing human values in their technical work. Indeed, at this time, only two communities were doing so, and both were working in well-defined contexts tied to the workplace.

The computer-supported cooperative work community (Galegher, Kraut, & Egido, 1990; Greif, 1988; Grudin, 1988) focused initially on the design and development of new technologies to help people collaborate effectively in the workplace—typically computer professionals working in relatively small groups and sometimes remotely. At the time, the values considered in computer-supported cooperative work systems were closely tied to group activities and workplace issues: cooperation, of course, but also such values as privacy, autonomy, ownership, security, and trust.

Finally, the participatory design community (Bjerknes & Bratteteig, 1995; Bødker, 1990; Ehn, 1988; Greenbaum & Kyng, 1991; Kyng & Mathiassen, 1997), comprised primarily of Scandinavian technologists and designers working (again, at the time) in an environment with strong labor unions and co-determination laws, developed a new approach to system design and development that fundamentally sought to empower workers' knowledge and a sense of work practice into the system design and development

process. In terms of value considerations, participatory design historically had deeply embedded within it a commitment to democratization of the workplace and human welfare as well as techniques to address unequal power relations within workplace settings.

Value sensitive design emerged out of this intellectual landscape in the early to mid-1990s in response to the perceived need for a broad-based design approach to account for human values and social context. In contrast to the research communities above, value sensitive design specifically targeted the design and development process to enable technologists and others to be proactive about engaging human values in the design process, and it took a broad perspective on human values, type of technology, and context of use.

Why the name "value sensitive design?" Labels matter. During this period, a wide range of variations on the label were considered. Following labels popular at the time (e.g., user-centered design, usable design, human-centered design), some options included value-centered design and value-based design. However, these were set aside as they seemed to imply that values would dominate other considerations in the technical design process. From the beginning, value sensitive design was conceptualized as an approach that would be engaged alongside of and intermingle with existing, well-functioning technical approaches. A label was sought to convey that relationship—to bring forward (be sensitive to) human values, but at the same time not to supplant the important technical efforts in their own right. Hence, the name: value sensitive design.

In its earliest published form, value sensitive design appeared hyphenated with the first letter of each word capitalized—"Value-Sensitive Design." As the term gained some purchase in the literature, the hyphen seemed unnecessary and awkward, and was dropped; hence, "Value Sensitive Design." Then, as the approach moved beyond a small group of initial researchers to be appropriated more broadly, any earlier sense of the term as a brand for a particular set of people's work was supplanted to reflect wider use, and the initial capital letters dropped except for titles or to introduce the approach; thus, "value sensitive design." Notably, similar changes in capitalization and use can be seen for the terms "participatory design" and "computer-supported cooperative work."

Bounding the Book

The value sensitive design literature is large and experiencing rapid growth. A Google Scholar search in January 2018 on the phrase "value sensitive design" returned over 3,500 works. A Google Scholar search on "value sensitive design" by year, from 2010 to 2016, returned 460 new works in 2010 and 935 new works in 2016, suggesting a growth trajectory. This is one reasonable, if imperfect, indicator of impact. In addition, the diffusion across fields appears to be significant, with publications ranging for example from workload management (Harbers & Neerincx, 2017) and digital journalism (Dörr & Hollnbuchner, 2017) to sustainable energy (Mok & Hyysalo, 2018).

Methodological development and innovation is rich within this body of work. Methods and theory engage, for example, transcultural and cross-cultural design (Alsheikh, Rode, & Lindley, 2011; Burmeister, 2013; Pereira & Baranauskas, 2015; Abokhodair & Vieweg, 2016); health informatics (Schikhof, Mulder, & Choenni, 2010; Huldtgren, Wiggers, & Jonker, 2014; Burmeister, 2016; Fitzpatrick, Huldtgren, Malmborg, Harley, & Ijsselsteijn, 2015; Novitzky et al., 2015; Pakrasi, Burmeister, Coppola, McCallum, & Loeb, 2015; Teipel et al., 2016; Grünloh, 2018); care robots in health settings (van Wynsberghe, 2013, 2015; Felzmann, Beyan, Ryan, & Beyan, 2016); empowerment and marginalization in crowd-work (Deng, Joshi, & Galliers, 2016); appropriation within action research (Weibert, Randall, & Wulf, 2017); embedding ethical and moral considerations throughout the software development lifecycle (Harbers, Detweiler, & Neerincx, 2015; Spiekermann, 2015; Ferrario et al., 2016); responsible innovation and value sensitive design (van den Hoven, 2013); and still other developments in varied application domains (e.g., Walldius, Sundblad, & Borning, 2005; Pommeranz, Detweiler, Wiggers, & Jonker, 2012; Shilton, 2012; van de Poel, 2013; Walldius & Lantz, 2013; Solomon, 2014; Stark & Tierney, 2014; JafariNaimi, Nathan, & Hargraves, 2015; Millar, 2016).

Furthermore, over the past 20 or more years, a large amount of work has been inspired by or developed in relation to value sensitive design. As a case in point, many different terms have been introduced to frame and describe work related to technology, values, and design, including: reflective design (Sengers, Boehner, David, & Kaye, 2005); value-centered human-computer interaction (Cockton, 2004); value-centered design (Cockton,

2005; Knight, 2008), later renamed worth-centered design (Cockton, 2006; Camara & Calvary, 2015); value-conscious design (Belman, Nissenbaum, Flanagan, & Diamond, 2011; Manders-Huits & Zimmer, 2009); values at play (Flanagan, Howe, & Nissenbaum, 2005); values for design (van den Hoven, 2005); values in design (Flanagan, Howe, & Nissenbaum, 2008; Knobel & Bowker, 2011); and, most recently, value-inspired design (Purao & Wu, 2013), ethical, value-based IT system design (Spiekermann, 2015), and values in computing (www.valuesincomputing.org). Simply put, this work largely seeks to move user-centered design, and more generally human-computer interaction, toward an even greater focus on human values—that is, what people believe to be important (Bannon, 2011; Harper, Rodden, Rogers, & Sellen, 2008).

Much additional literature is relevant to value sensitive design. Early interest in technology, values, and design can be found, for example, in the work of Mumford (1934), Wiener (1954/1985, 1985), Papanek (1971), and Kling (1980). Fields such as computer ethics, information systems, media studies, and science and technology studies are also relevant (for a review, see Snyder, Shilton, & Anderson, 2016). Numerous perspectives on design, moreover, are highly relevant to the theory, method, and practice of value sensitive design, including ability-based design (Wobbrock, Gajos, Kane, & Vanderheiden, 2018); adversarial design (DiSalvo, 2012); collapse informatics (Tomlinson et al., 2013); contestational design and surreptitious communication design (Hirsch, 2008; Hirsch, 2016); critical design (Malpass, 2013; Bardzell, Bardzell, & Stolterman, 2014); feminist design (Bardzell, 2010; Rode, 2011); empowered design (Marsden, 2008); ludic design (Blythe et al., 2010); positive computing (Calvo & Peters, 2014); postcolonial computing (Irani, Vertesi, Dourish, Philip, & Grinter, 2010; Philip, Irani, & Dourish, 2012); privacy by design (Koops & Leenes, 2014); sustainable interaction design (Blevis, 2007; Nathan, 2012); and systems for human benefit (Venable, Pries-Heje, Bunker, & Russo, 2011), among many others. In addition, participatory design has continued to evolve vigorously (Ehn, 2008; Binder et al., 2011; Vines, Clarke, Wright, McCarthy, & Olivier, 2013), and its older and more recent contributions to the literature are highly relevant to value sensitive design.

It is beyond the scope of this book to pursue a thorough analysis of this large literature (for a review, see Shilton, 2018). Instead, our goal is to bring together in one place a current articulation of value sensitive design.

Beginning with theory, we discuss the underpinnings of value sensitive design. Then, we bring together 17 value sensitive design methods, showing how theory and method go hand in hand. We discuss each method and explore how it can be used to address values in the research and design processes. The book also includes reports on applications from 10 technical domains, each authored by leading researchers in that domain. We conclude with a discussion of several published critiques, with an assessment of the robustness of the current state of value sensitive design, and lastly with some reflections for furthering and deepening value sensitive design. We end with a commitment to progress, not perfection.

do computers, spaceships, and nuclear power plants. Some might wish to argue that even stone axe heads, wooden ploughs, and paring knives belong here as well, since they, too, involve the application of at least some sort of rudimentary scientific knowledge. In any case, there is clearly a continuum. A sharp stone used for shaping wood is a tool, as is, presumably, a mechanical hand drill. What about an electric drill? A computer-controlled milling machine? We would prefer not to belabor this point; the difference is at most one of degree. To tools and technology, we then add *infrastructure*—the basic physical and organizational structures and facilities needed for the operation of a society or enterprise. This includes the buildings, roads, energy sources, and other structures that enable complex societal activities such as communication, transportation, and information flow. Taken together, tools, technology, and infrastructure comprise what some might term a technological system. When speaking of one—tool, technology, or infrastructure—it is nearly impossible not to speak of the others. For purposes of this book, we will use the term technology as a shorthand to refer to all three and their interdependencies.

We tend to view technology in terms of artifacts. But it is also possible to view policy in this light—particularly as some combination of tool and infrastructure. After all, policy also shapes, albeit through law and regulation, human activity. For purposes of value sensitive design, we currently consider policy to be a form of technology.

The design and diffusion of technologies typically involve diverse actors and relationships. One kind of relationship exists between the actors in the design and use contexts (Albrechtslund, 2007). Here, clients, co-designers, designers, field-testers, inventors, and so forth imagine and develop technology that, later, is appropriated and integrated into society by other kinds

of actors. Limited in part by the human ability to foresee, designers simply cannot anticipate the many ways a technology will be adopted, appropriated, or, in the extreme, used subversively; nor can designers fully anticipate the consequences of a technology on social life. That said, we shall see that value sensitive design offers methods (e.g., direct and indirect stakeholder analysis, value scenarios, Envisioning Cards) that position designers to envision the many actors and relationships that emerge through the technology design process.

At its core, value sensitive design is technology agnostic. That is, in principle the theory, method, and practice of value sensitive design are not tied to any specific technology. They could just as well be applied to the development of brain-machine interfaces as to technologies for agriculture, energy, or water treatment. That said, the vast majority of early work in value sensitive design concerned information technology. For example, projects engaged with artificial intelligence and autonomous agents (Friedman & Kahn, 1992; Friedman & Nissenbaum, 1997); computer algorithms (Friedman & Nissenbaum, 1996); web browser security (Friedman, Howe, & Felten, 2002); human-robot interaction (Kahn et al., 2007; Kahn, Friedman, Perez-Granados, & Freier, 2006); location-aware mobile devices (Czeskis et al., 2010; Friedman, Smith, Kahn, Consolvo, & Selawski, 2006c); large-scale urban simulation (Borning, Friedman, Davis, & Lin, 2005; Davis et al., 2006); and groupware knowledge systems (Miller, Friedman, Jancke, & Gill, 2007). With the emergence of the Internet of Things, we see an increased blurring between information systems, on the one hand, and physical systems on the other. Here value sensitive design has been applied to "things" such as wireless implantable cardiac devices (Denning et al., 2010) and public transportation information systems (Watkins, Ferris, Malinovskiy, & Borning, 2013b). More recently, value sensitive design has been applied to non-information technologies and processes such as wastewater treatment (de Kreuk, van de Poel, Zwart, & van Loosdrecht, 2010; van de Poel, Zwart, Brumsen, & van Mil, 2005) and wind turbine parks (Oosterlaken, 2015). It remains an open practical question how (if at all) the theory and method of value sensitive design developed primarily with information technologies will need to be adapted or extended to account for human values in the design process of other non-information technology.

Human Values

We turn now to explicate how the term "human values" has been understood within value sensitive design.

In 1997, in the introduction to her edited book on *Human Values and the Design of Computer Technology*, Friedman provided a working definition of human values as follows:

> In some sense, we can say that any human activity reflects human values. I drink tea instead of soda. I recently attended a Cezanne exhibit instead of a ball game. I have personal values. We all do. But these are not the type of human values which this volume takes up. Rather, this volume is principally concerned with values that deal with human welfare and justice. (p. 3)

This framing placed an emphasis on human welfare and justice—what some might call moral values. Moreover, Friedman pointed to ways in which culture and context inform people's understandings and experiences of both welfare and justice (e.g., in Western societies, most people believe that if a person chooses to wear a bathrobe that is a personal choice; but if a person wears a bathrobe to a dignitary's funeral, many people would judge that event not only as a conventional violation, but a moral violation as well—a sign of disrespect). Following Turiel (1983), Friedman further drew on the psychological literature to distinguish among three broad domains of social knowledge: moral, conventional, and personal. Here, the moral domain refers to prescriptive judgments that people justify on the basis of considerations of justice, fairness, rights, or human welfare. The conventional domain, in contrast, refers to judgments concerning behavioral uniformities that help to promote the smooth functioning of

social interactions, and the personal domain to judgments seen as under the jurisdiction of the self. Friedman concluded her discussion with the rationale for focusing on human values rather than more narrowly on only moral ones.

> The human values addressed in this volume principally refer to moral values. But I prefer to use the broader term "human values" instead of simply moral values to highlight the complexity of social life, and to provide a basis for analyses wherein personal and conventional values can become morally implicated. (p. 5)

Thus, from its earliest conceptions, value sensitive design sought to emphasize moral and ethical values, but to do so within the complexity of social life and with recognition for how culture and context implicate people's understanding and experience of harms and injustice.

A decade later, Friedman et al. (2006a) expanded the discussion of values within value sensitive design to emphasize *what is important to people in their lives*. Within this broad and gentle conceptualization of "value," they also reaffirmed a focus on ethical and moral values, while at the same time calling attention to the long and ongoing contentious history of framing human values. They wrote:

> In a narrow sense, the word "value" refers simply to the economic worth of an object. For example, the value of a computer could be said to be two thousand dollars. However, in the work described here [value sensitive design], we use a broader meaning of the term wherein a value refers to what a person or group of people consider important in life.[1] In this sense, people find many things of value, both lofty and mundane: their children, friendship, morning tea, education, art, a walk in the woods, nice manners, good science, a wise leader, clean air.
>
> This broader framing of values has a long history. Since the time of Plato, for example, the content of value-oriented discourse has ranged widely, emphasizing "the good, the end, the right, obligation, virtue, moral judgment, aesthetic judgment, the beautiful, truth, and validity" [Frankena, 1972, p. 229]. Sometimes ethics has been subsumed within a theory of values, and other times conversely, with ethical values viewed as just one component of ethics more generally. Either way, it is usually agreed [Moore, 1903/1978] that values should not be conflated with facts (the "fact/value distinction") especially insofar as facts do not logically entail value. In other words, "is" does not imply "ought" (the naturalistic fallacy). In this way, values cannot be motivated only by an empirical account of the external world, but depend substantively on the interests and desires of human beings within a cultural milieu.

[1]The Oxford English Dictionary definition of this sense of value is: "the principles or standards of a person or society, the personal or societal judgment of what is valuable and important in life." (Friedman et al., 2006a, p. 349)

Since 2006, reflection on completed and ongoing work has led to this current working definition of "value" within value sensitive design: *what is important to people in their lives, with a focus on ethics and morality*.

Over the years, critiques have surfaced about how value sensitive design has conceptualized "human value." They take primarily two forms, largely reflecting debates in moral philosophy and the social sciences. For those who seek normative positions (these tend to be moral philosophers), the current formulation of human values within value sensitive design is underspecified (see, for example, Albrechtslund, 2007; Manders-Huits, 2011; van de Poel, 2009). Conversely, for those with leanings toward culture-, community-, or individual-specific positions—or who believe that value sensitive design should accommodate researchers with these views—the current formulation of human values within value sensitive design may imply too great an underlying universality (Borning & Muller, 2012; Le Dantec, Poole, & Wyche, 2009). Granted, both types of dissatisfactions are understandable. After all, the working definition of "human values," while retaining moral and ethical sensibilities, does not spell out *what* is important to people in their lives. That work is left to individual researchers and designers as they move forward on specific projects. For example, some might wish to draw on ethical principles from Buddhism, others from a care ethics, still others from a consequentialist or deontic position. At the same time, the working definition is intended to hold a commitment to moral and ethical sensibilities. Thus, it is incumbent on those working from a culture-, community-, or individual-specific perspective to establish the basis for and bring forward the ethical and moral dimensions within that perspective.

Given the complexity of social life, the technological development process, the commitments of designers, and the unresolved nature of debates on morality, this working definition provides an appropriate balance. That is, it positions value sensitive design to engage with important moral and ethical considerations in light of technology design and legitimates other considerations that are important to people in their lives; it leaves the explication of both to the details of any particular design project. For example, van Wynsberghe (2013) engages value sensitive design from a perspective of care ethics, Cummings (2006) from that of a just war. In these respective perspectives, both authors pay careful attention to the moral and ethical

was intended to mean human values more generally (and not the specific human values in the list presented here); the original wording was unfortunately ambiguous.

They continued:

> Two caveats: Not all of the values that we review are fundamentally distinct from one another. Nonetheless, each value has its own language and conceptualizations within their respective fields, and thus warrants separate treatment here. Second, this list is not comprehensive. Perhaps no list could be, at least within the confines of a chapter. Peacefulness, compassion, love, warmth, creativity, humor, originality, vision, friendship, cooperation, collaboration, purposefulness, devotion, diplomacy, kindness, musicality, harmony—the list of other possible values could get very long very fast. Our particular list comprises many of the traditional values that hinge on the deontological and consequentialist moral orientations: Human welfare, ownership and property, privacy, freedom from bias, universal usability, trust, autonomy, informed consent, and accountability. In addition, we have chosen several nontraditional values within the [human-computer interaction] … community: Identity, calmness, and environmental sustainability. Our goal here is not only to point to important areas of future inquiry, but also to illustrate how an overarching framework for human values and ethics in design can move one quickly and substantively into new territory. (p. 1187)

A few years later, Friedman et al. (2006a) revisited this list, with an emphasis on human values with ethical import that were often implicated in system design. We reproduce the table with that list of values and working definitions, omitting the column of references (see table 2.1). The pragmatic concerns remain: how to avoid reification of a certain set of values or world views, while at the same time positioning those bringing value sensitive design into their research and design processes to build upon each others' prior work. The tensions here may best be addressed through practice (see Borning & Muller [2012] for some suggestions).

Beyond human values. One last reflection on human values—what *people* consider important in their lives. This framing privileges the perspectives and values of human beings. After all, value sensitive design concerns the design process of technology that is carried out by human beings. Yet the technologies we design and build reach far beyond human beings to implicate other nonhuman entities. Here are four for consideration: nonhuman species, superorganisms, the Earth, and social robots. In sorting out how to account for nonhumans in value sensitive design, both homocentric (valuing nonhumans because of what they offer human beings) and biocentric