Value Sensitive Design and Information Systems

BATYA FRIEDMAN, PETER H. KAHN, JR., AND ALAN BORNING
University of Washington

Forthcoming in P. Zhang & D. Galletta (Eds.), *Human-Computer Interaction in Management Information Systems: Foundations*. M.E. Sharpe, Inc: NY.

Value Sensitive Design is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process. It employs an integrative and iterative tripartite methodology, consisting of conceptual, empirical, and technical investigations. We explicate Value Sensitive Design by drawing on three case studies. The first study concerns information and control of web browser cookies, implicating the value of informed consent. The second study concerns using high-definition plasma displays in an office environment to provide a "window" to the outside world, implicating the values of physical and psychological well-being and privacy in public spaces. The third study concerns an integrated land use, transportation, and environmental simulation system to support public deliberation and debate on major land use and transportation decisions, implicating the values of fairness, accountability, and support for the democratic process, as well as a highly diverse range of values that might be held by different stakeholders, such as environmental sustainability, opportunities for business expansion, or walkable neighborhoods. We conclude with direct and practical suggestions for how to engage in Value Sensitive Design.

---

1. INTRODUCTION

There is a longstanding interest in designing information and computational systems that support enduring human values. Researchers have focused, for example, on the value of *privacy* [Ackerman and Cranor 1999; Agre and Rotenberg 1998; Fuchs 1999; Jancke et al. 2001; Palen and Grudin 2003; Tang 1997], *ownership* and *property* [Lipinski and Britz 2000], *physical welfare* [Leveson 1991], *freedom from bias* [Friedman and Nissenbaum 1996], *universal usability* [Shneiderman 1999, 2000; Thomas 1997], *autonomy* [Suchman 1994; Winograd 1994], *informed consent* [Millett et al. 2001], and *trust* [Fogg and Tseng 1999; Palen and Grudin 2003; Riegelsberger and Sasse 2002; Rocco 1998; Zheng et al. 2001]. Still, there is a need for an overarching theoretical and methodological framework with which to handle the value dimensions of design work.

Value Sensitive Design is one effort to provide such a framework (e.g., Friedman [1997a], Friedman and Kahn [2003], Friedman and Nissenbaum [1996], Hagman, Hendrickson, and Whitty [2003], Nissenbaum [1998], Tang [1997], and Thomas [1997]). Our goal in this paper is to provide an account of Value Sensitive Design, with enough detail for other researchers and designers to critically examine and systematically build on this approach.

We begin by sketching the key features of Value Sensitive Design, and then describe its integrative tripartite methodology, which involves conceptual, empirical, and technical investigations, employed iteratively. Then we explicate Value Sensitive Design by drawing on three case studies. One involves cookies and informed consent in web browsers; the second involves HDTV display technology in an office environment; the third involves user interactions and interface for an integrated land use, transportation, and environmental simulation. We conclude with direct and practical suggestions for how to engage in Value Sensitive Design.

2. WHAT IS VALUE SENSITIVE DESIGN?

Value Sensitive Design is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process.

2.1. What is a Value?

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, 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. In Table 1 in Section 6.8, we provide a list of human values with ethical import that are often implicated in system design, along with working definitions and references to the literature.

2.2. Related Approaches to Values and System Design

In the 1950’s, during the early periods of computerization, cyberneticist Norbert Wiener [1953/1985] argued that technology could help make us better human beings, and create a more just society. But for it to do so, he argued, we have to take control of the technology. We have to reject the “worshiping [of] the new gadgets which are our own creation as if they were our masters” (p. 678). Similarly, a few decades later, computer scientist Joseph Weizenbaum [1972] wrote:

> What is wrong, I think, is that we have permitted technological metaphors…and technique itself to so thoroughly pervade our thought processes that we have finally abdicated to technology the very duty to formulate questions…Where a simple man might ask: “Do we need these things?”, technology asks “what electronic wizardry will make them safe?” Where a simple man will ask “is it good?”, technology asks “will it work?” (pp. 611-612)

More recently, supporting human values through system design has emerged within at least four important approaches. *Computer Ethics* advances our understanding of key values that lie at the intersection of computer technology and human lives, e.g., Bynum [1985], Johnson and Miller [1997], and Nissenbaum [1999]. *Social Informatics* has been successful in providing socio-technical analyses of deployed technologies, e.g., Kling, Rosenbaum, and Hert [1998], Kling and Star [1998], and Sawyer and Rosenbaum [2000].

---

[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 judgement of what is valuable and important in life.” [Simpson and Weiner 1989]

Computer Supported Cooperative Work (CSCW) has been successful in the design of new technologies to help people collaborate effectively in the workplace, e.g., Fuchs [1999], Galegher, Kraut, and Egido [1990], Olson and Teasley [1996], and Grudin [1988]. Finally, *Participatory Design* substantively embeds democratic values into its practice, e.g., Bjerknes & Bratteteig [1995], Bødker [1990], Ehn [1989], Greenbaum and Kyng [1991], and Kyng and Mathiassen [1997]. (See Friedman and Kahn [2003] for a review of each of these approaches.)

3. THE TRIPARTITE METHODLOGY: CONCEPTUAL, EMPIRICAL, AND TECHNICAL INVESTIGATIONS

Think of an oil painting by Monet or Cézanne. From a distance it looks whole; but up close you can see many layers of paint upon paint. Some paints have been applied with careful brushstrokes, others perhaps energetically with a palate knife or fingertips, conveying outlines or regions of color. The diverse techniques are employed one on top of the other, repeatedly, and in response to what has been laid down earlier. Together they create an artifact that could not have been generated by a single technique in isolation of the others. So, too, with Value Sensitive Design. An artifact (e.g., system design) emerges through iterations upon a process that is more than the sum of its parts. Nonetheless, the parts provide us with a good place to start. Value Sensitive Design builds on an iterative methodology that integrates conceptual, empirical, and technical investigations; thus, as a step toward conveying Value Sensitive Design, we describe each investigation separately.

3.1 Conceptual Investigations

Who are the direct and indirect stakeholders affected by the design at hand? How are both classes of stakeholders affected? What values are implicated? How should we engage in trade-offs among competing values in the design, implementation, and use of information systems (e.g., autonomy vs. security, or anonymity vs. trust)? Should moral values (e.g., a right to privacy) have greater weight than, or even trump, non-moral values (e.g., aesthetic preferences)? Value Sensitive Design takes up these questions under the rubric of conceptual investigations.

In addition, careful working conceptualizations of specific values clarify fundamental issues raised by the project at hand, and provide a basis for comparing results across research teams. For example, in their analysis of trust in online system design, Friedman, Kahn, and Howe [2000], drawing on Baier [1986], first offer a philosophically informed working conceptualization of trust. They propose that people trust when they are vulnerable to harm from others, yet believe those others would not harm them even though they could. In turn, trust depends on people's ability to make three types of assessments. One is about the harms they might incur. The second is about the good will others possess toward them that would keep those others from doing them harm. The third involves whether or not harms that do occur lie outside the parameters of the trust relationship. From such conceptualizations, Friedman et al. were able to define clearly what they meant by trust online. This definition is in some cases different from what other researchers have meant by the term – for example, the Computer Science and Telecommunications Board, in their thoughtful publication *Trust in Cyberspace* [Schneider 1999], adopted the terms "trust" and "trustworthy" to describe systems that perform as expected along the dimensions of correctness, security, reliability, safety, and survivability. Such a definition, which equates "trust" with expectations for machine performance, differs markedly from one that says trust is fundamentally a relationship between people (sometimes mediated by machines).

3.2 Empirical Investigations

Conceptual investigations can only go so far. Depending on the questions at hand, many analyses will need to be informed by empirical investigations of the human context in which the technical artifact is situated. Empirical investigations are also often needed to evaluate the success of a particular design. Empirical investigations can be applied to any human activity that can be observed, measured, or documented. Thus, the entire range of quantitative and qualitative methods used in social science research is potentially applicable here, including observations, interviews, surveys, experimental manipulations, collection of relevant documents, and measurements of user behavior and human physiology.

Empirical investigations can focus, for example, on questions such as: How do stakeholders apprehend individual values in the interactive context? How do they prioritize competing values in design trade-offs? How do they prioritize individual values and usability considerations? Are there differences between espoused practice (what people say) compared with actual practice (what people do)? Moreover, because the development of new technologies affects groups as well as individuals, questions emerge of how organizations appropriate value considerations in the design process. For example, regarding value considerations, what are organizations' motivations, methods of training and dissemination, reward structures, and economic incentives?

3.3 Technical Investigations

As discussed in Section 2.3 (Value Sensitive Design's Constellation of Features), Value Sensitive Design adopts the position that technologies in general, and information and computer technologies in particular, provide value suitabilities that follow from properties of the technology. That is, a given technology is more suitable for certain activities and more readily supports certain values while rendering other activities and values more difficult to realize.

In one form, technical investigations focus on how existing technological properties and underlying mechanisms support or hinder human values. For example, some video-based collaborative work systems provide blurred views of office settings, while other systems provide clear images that reveal detailed information about who is present and what they are doing. Thus the two designs differentially adjudicate the value trade-off between an individual's *privacy* and the group's *awareness* of individual members' presence and activities.

In the second form, technical investigations involve the proactive design of systems to support values identified in the conceptual investigation. For example, Fuchs [1999] developed a notification service for a collaborative work system in which the underlying technical mechanisms implement a value hierarchy whereby an individual's desire for privacy overrides other group members' desires for awareness.

At times, technical investigations – particularly of the first form – may seem similar to empirical investigations insofar as both involve technological and empirical activity. However, they differ markedly on their unit of analysis. Technical investigations focus on the technology itself. Empirical investigations focus on the individuals, groups, or larger social systems that configure, use, or are otherwise affected by the technology.

4. VALUE SENSITIVE DESIGN IN PRACTICE: THREE CASE STUDIES

To illustrate Value Sensitive Design's integrative and iterative tripartite methodology, we draw on three case studies with real world applications, one completed and two under way. Each case study represents a unique design space.

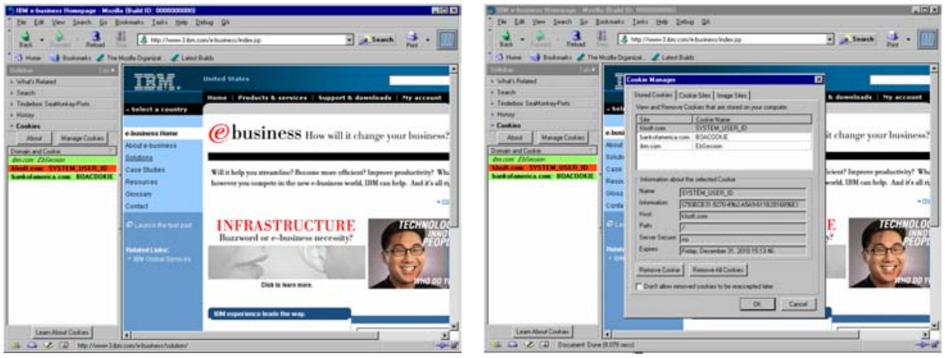
4.1 Cookies and Informed Consent in Web Browsers

Informed consent provides a critical protection for privacy, and supports other human values such as autonomy and trust. Yet currently there is a mismatch between industry practice and the public's interest. According to a recent report from the Federal Trade Commission (2000), for example, 59% of Web sites that collect personal identifying information neither inform Internet users that they are collecting such information nor seek the user's consent. Yet, according to a Harris poll (2000), 88% of users want sites to garner their consent in such situations.

Against this backdrop, Friedman, Felten, and their colleagues [Friedman et al. 2002; Friedman et al. 2000; Millett et al. 2001] sought to design web-based interactions that support informed consent in a web browser through the development of new technical mechanisms for cookie management. This project was an early proof-of-concept project for Value Sensitive Design, which we use here to illustrate several key features of the methodology.

4.1.1 Conceptualizing the Value. One part of a conceptual investigation entails a philosophically informed analysis of the central value constructs. Accordingly, Friedman et al. began their project with a conceptual investigation of informed consent itself. They drew on diverse literature, such as the Belmont Report, which delineates ethical principles and guidelines for the protection of human subjects [Belmont Report 1978; Faden and Beauchamp 1986], to develop criteria for informed consent in online interactions. In brief, the idea of "informed" encompasses disclosure and comprehension. *Disclosure* refers to providing accurate information about the benefits and harms that might reasonably be expected from the action under consideration. *Comprehension* refers to the individual's accurate interpretation of *what* is being disclosed. In turn, the idea of "consent" encompasses voluntariness, comprehension, and agreement. *Voluntariness* refers to ensuring that the action is not controlled or coerced. *Competence* refers to possessing the mental, emotional and physical capabilities needed to be capable of giving informed consent. *Agreement* refers to a reasonably clear opportunity to accept or decline to participate. Moreover, agreement should be ongoing, that is, the individual should be able to withdraw from the interaction at any time. See Friedman, Millet, and Felten [2000] for an expanded discussion of these five criteria.

4.1.2 Using a Conceptual Investigation to Analyze Existing Technical Mechanisms. With a conceptualization for informed consent online in hand, Friedman et al. conducted a retrospective analysis (one form of a technical investigation) of how the cookie and web-browser technology embedded in Netscape Navigator and Internet Explorer changed with respect to informed consent over a 5-year period, beginning in 1995. Specifically, they used the criteria of disclosure, comprehension, voluntariness, competence, and agreement to evaluate how well each browser in each stage of its development supported the users' experience of informed consent. Through this retrospective analysis, they found that while cookie technology had improved over time regarding informed consent (e.g., increased visibility of cookies, increased options for accepting or declining cookies, and access to information about cookie content), as of 1999 some startling problems remained. For example: (a) While browsers disclosed to users some information about cookies, they still did not disclose the right sort of information – that is, information about the potential harms and benefits from setting a particular cookie. (b) In Internet Explorer, the burden to accept or decline all third party cookies still fell to the user, placing undue burden on the user to decline each third party cookie one at a time. (c) Users' out-of-the-box experience of cookies (i.e., the default setting) was no different in 1999 than it was in 1995: to accept all cookies. That is, the novice user installed a browser that accepted all cookies and disclosed nothing about that activity to the user. (d) Neither browser alerted a user when a site wished to use a cookie and for what purpose, as opposed to when a site wished to store a cookie.



(a) Peripheral awareness mechanism.

(b) Just-in-time cookie management tool.

Figure 1. Screen shot (a) of the Mozilla implementation shows the peripheral awareness of cookies interface (at the left) in the context of browsing the web. Each time a cookie is set, a color-coded entry for that cookie appears in the sidebar. Third party cookies are red; others are green. At the user's discretion, he or she can click on any entry to bring up the Mozilla cookie manager for that cookie. Screen shot (b) after the user has clicked on an entry to bring up the just-in-time cookie management tool (in the center) for a particular cookie.

4.1.3 The Iteration and Integration of Conceptual, Technical, and Empirical Investigations. Based on the results from these conceptual and technical investigations, Friedman et al. then iteratively used the results to guide a second technical investigation: a redesign of the Mozilla browser (the open-source code for Netscape Navigator). Specifically, they developed three new types of mechanisms: (a) peripheral awareness of cookies; (b) just-in-time information about individual cookies and cookies in general; and (c) just-in-time management of cookies (see Figure 1). In the process of their technical work, Friedman et al. conducted formative evaluations (empirical investigations) which led to a further design criterion, *minimal distraction*, which refers to meeting the above criteria for informed consent without unduly diverting the user from the task at hand. Two situations are of concern here. First, if users are overwhelmed with queries to consent to participate in events with minor benefits and risks, they may become numbed to the informed consent process by the time participation in an event with significant benefits and risks is at hand. Thus, the user's participation in that event may not receive the careful attention that is warranted. Second, if the overall distraction to obtain informed consent becomes so great as to be perceived to be an intolerable nuisance, users are likely to disengage from the informed consent process in its entirety and accept or decline participation by rote. Thus undue distraction can single-handedly undermine informed consent. In this way, the iterative results of the above empirical investigations not only shaped and then validated the technical work, but impacted the initial conceptual investigation by adding to the model of informed consent the criterion of minimal distraction.

Thus, this project illustrates the iterative and integrative nature of Value Sensitive Design, and provides a proof-of-concept for Value Sensitive Design in the context of mainstream Internet software.

4.2 Room with a View: Using Plasma Displays in Interior Offices

> Janice is in her office, writing a report. She's trying to conceptualize the report's higher-level structure, but her ideas won't quite take form. Then she looks up from her desk and rests her eyes on the fountain and plaza area outside her building. She notices the water bursting upward, and that a small group of people are gathering by the water's edge. She rests her eyes on the surrounding pool of calm water. Her eyes then lift toward the clouds and the streaking sunshine. Twenty seconds later she returns to her writing task at hand, slightly refreshed, and with an idea taking shape.

What's particularly novel about this workplace scenario is that Janice works in an interior office. Instead of a real window looking out onto the plaza, Janice has a large screen video plasma display that continuously displays the local outdoor scene in real-time. Realistic? Beneficial? This design space is currently being researched by Kahn, Friedman, and their colleagues, using the framework of Value Sensitive Design.

In Kahn et al.'s initial conceptual investigation of this design space, they drew on the psychological literature that suggests that interaction with real nature can garner physiological and psychological benefits. For example, in one study, Ulrich [1984] found that post-operative recovery improved when patients were assigned to a room with a view of a natural setting (a small stand of deciduous trees) versus a view of a brown brick wall. More generally, studies have shown that even minimal connection with nature – such as looking at a natural landscape – can reduce immediate and long-term stress, reduce sickness of prisoners, and calm patients before and during surgery. (See Beck and Katcher [1996], Kahn [1999], and Ulrich [1993] for reviews.) Thus Kahn et al. hypothesized that an "augmented window" of nature could render benefits in a work environment in terms of the human values of physical health, emotional well-being, and creativity.

To investigate this question in a laboratory context, Kahn et al. are comparing the short-term benefits of working in an office with a view out the window of a beautiful nature scene versus an identical view (in real time) shown on a large video plasma display that covers the window in the same office (Figure 2a). In this latter condition, they employed a High Definition TV (HDTV) camera (Figure 2b) to capture real-time local images. The control condition involved a blank covering over the window. Their measures entailed (a) physiological data (heart rate), (b) performance data (on cognitive and creativity tasks), (c) video data that captured each subject's eye gaze on a second-by-second level, and time synchronized with the physiological equipment, so that analyses can determine whether physiological benefits accrued immediately following an eye gaze onto the plasma screen, and (d) social-cognitive data (based on a 50-minute interview with each subject at the conclusion of the experimental condition wherein they garnered each subject's reasoned perspective on the experience). Data analysis is in progress. However, preliminary results are showing the following trends. First, participants looked out the plasma screen just as frequently as they did the real window, and more frequently than they stared at the blank wall. In this sense, the plasma-display window was functioning like a real window. But, when participants gazed for 30 seconds or more, the real window provided greater physiological recovery from low-level stress as compared to the plasma display window.

(a) "The Watcher" (b) The HDTV Camera (c) "The Watched"

Figure 2. Plasma Display Technology Studies

From the standpoint of illustrating Value Sensitive Design, we would like to emphasize five ideas.

4.2.1. Multiple Empirical Methods. Under the rubric of empirical investigations, Value Sensitive Design supports and encourages multiple empirical methods to be used in concert to address the question at hand. As noted above, for example, this study employed physiological data (heart rate), two types of performance data (on cognitive and creativity tasks), behavioral data (eye gaze), and reasoning data (the social-cognitive interview). From a value-oriented perspective, multiple psychological measures increase the veracity of most accounts of technology in use.

4.2.2. Direct and Indirect Stakeholders. In their initial conceptual investigation of the values implicated in this study, Kahn et al. sought to identify not only direct but also indirect stakeholders affected by such display technology. At that early point, it became clear to the researchers that an important class of indirect stakeholders (and their respective values) needed to be included: namely, the individuals who, by virtue of walking through the fountain scene, unknowingly had their images displayed on the video plasma display in the "inside" office (Figure 2c). In other words, if this application of projection technology were to come into widespread use (as web cams and surveillance cameras have begun to) then it would potentially encroach on the privacy of individuals in public spaces – an issue that has been receiving increasing attention in the field of computer ethics and public discourse [Nissenbaum 1998]. Thus, in addition to the experimental laboratory study, Kahn et al. initiated two additional but complementary empirical investigations with indirect stakeholders: (a) a survey of 750 people walking through the public plaza, and (b) in-depth social cognitive interviews with 30 individuals walking through the public plaza [Friedman, Kahn, and Hagman 2004]. Both investigations focused on indirect stakeholders' judgments of privacy in public space, and in particular having their real-time images captured and displayed on plasma screens in nearby and distant offices. The importance of such indirect stakeholder investigations is being borne out by the results. For example, significant gender differences were found in their survey data: more women than men expressed concern about the invasion of privacy through web cameras in public places. This finding held whether their image was to be displayed locally or in another city (Tokyo), or viewed by one person, thousands, or millions. One implication of this finding is that future technical designs and implementations of such display technologies need to be responsive to ways in which men and women might perceive potential harms differently.

4.2.3. Coordinated Empirical Investigations. Once Kahn et al. identified an important group of indirect stakeholders, and decided to undertake empirical investigations with this group, they then coordinated these empirical investigations with the initial (direct stakeholder) study. Specifically, a subset of identical questions were asked of both the direct stakeholders ("The Watchers") and indirect stakeholders ("The Watched"). Results show some interesting differences. For example, more men in The Watched condition expressed concerns about that people's images might be displayed locally, nationally, or internationally than men in The Plasma Display Watcher condition. No differences were found between women in The Watcher Plasma Display Condition and women in the Watched condition. Thus, the Value Sensitive Design methodology helps to bring to the forefront values that matter not only to the direct stakeholders of a technology (such as physical health, emotional well-being, and creativity), but to the indirect stakeholders (such as privacy, informed consent, trust, and physical safety). Moreover, from the standpoint of Value Sensitive Design, the above study highlights how investigations of indirect stakeholders can be woven into the core structure of the experimental design with direct stakeholders.

4.2.4. Multiplicity of and Potential Conflicts among Human Values. Value Sensitive Design can help researchers uncover the multiplicity of and potential conflicts among human values implicated in technological implementations. In the above design space, for example, values of physical health, emotional well-being, and creativity appear to partially conflict with other values of privacy, civil rights, trust, and security.

4.2.5. Technical Investigations. Conceptual and empirical investigations can help to shape future technological investigations, particularly in terms of how nature (as a source of information) can be embedded in the design of display technologies to further human well-being. One obvious design space involves buildings. For example, if Kahn et al.'s empirical results continue to emerge in line with their initial results, then one possible design guideline is as follows: we need to design buildings with nature in mind, and within view. In other words, we cannot with psychological impunity digitize nature and display the digitized version as a substitute for the real thing (and worse, then destroy the original). At the same time, it is possible that technological representations of nature can garner some psychological benefits, especially when (as in an inside office) direct access to nature is otherwise unavailable. Other less obvious design spaces involve, for example, airplanes. In recent discussions with Boeing Corporation, for example, we were told that for economic reasons engineers might like to construct airplanes without passenger windows. After all, windows cost more to build and decrease fuel efficiency. At stake, however, is the importance of windows in the human experience of flying.

In short, this case study highlights how Value Sensitive Design can help researchers employ multiple psychological methods, across several studies, with direct and indirect stakeholders, to investigate (and ultimately support) a multiplicity of human values impacted by deploying a cutting-edge information technology.

4.3 UrbanSim: Integrated Land Use, Transportation, and Environmental Simulation

In many regions in the United States (and globally), there is increasing concern about pollution, traffic jams, resource consumption, loss of open space, loss of coherent community, lack of sustainability, and unchecked sprawl. Elected officials, planners, and citizens in urban areas grapple with these difficult issues as they develop and evaluate alternatives for such decisions as building a new rail line or freeway, establishing an urban growth boundary, or changing incentives or taxes. These decisions interact in complex ways, and, in particular, transportation and land use decisions interact strongly with each other. There are both legal and common sense reasons to try to understand the long-term consequences of these interactions and decisions. Unfortunately, the need for this understanding far outstrips the capability of the analytic tools used in current practice.

In response to this need, Waddell, Borning, and their colleagues have been developing UrbanSim, a large simulation package for predicting patterns of urban development for periods of twenty years or more, under different possible scenarios [Waddell 2002; Noth et al. 2003; Waddell et al. 2003]. Its primary purpose is to provide urban planners and other stakeholders with tools to aid in more informed decision-making, with a secondary goal to support further democratization of the planning process. When provided with different scenarios – packages of possible policies and investments – UrbanSim models the resulting patterns of urban growth and redevelopment, of transportation usage, and of resource consumption and other environmental impacts.

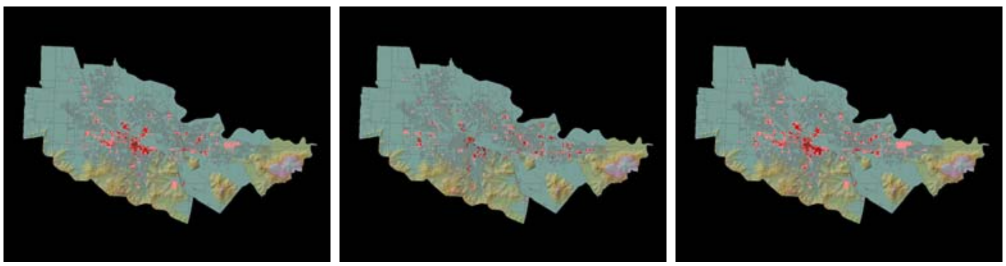


Figure 3. Results from UrbanSim for Eugene/Springfield, Oregon, forecasting land use patterns over a 14-year period. These results arise from the simulated interactions among demographic change, economic change, real estate development, transportation, and other actors and processes in the urban environment. Map (a) shows the employment density in 1980 (number of jobs located in each 150x150 meter grid cell). Darker red indicates higher density. Map (b) shows the predicted change from 1980 to 1994 (where darker red indicates a greater change), and map (c) the predicted employment density in 1994. In a historical validation of the model, this result was then compared with the actual 1994 employment, with a 0.917 correlation over a 1-cell radius.

To date, UrbanSim has been applied in the metropolitan regions around Eugene/Springfield, Oregon (Figure 3), Honolulu, Hawaii, Salt Lake City, Utah, and Houston, Texas, with application to the Puget Sound region in Washington State under way. UrbanSim is undergoing significant redevelopment and extension in terms of its underlying architecture, interface, and social goals. Under the direction of Borning, Friedman, and Kahn, Value Sensitive Design is playing a central role in this endeavor.

UrbanSim illustrates important aspects of Value Sensitive Design in addition to those described in the previous two case studies:

4.3.1 Distinguishing Explicitly Supported Values from Stakeholder Values. In their conceptual investigations, Borning et al. distinguished between explicitly supported values (i.e., ones that they explicitly want to embed in the simulation) and stakeholder values (i.e., ones that are important to some but not necessarily all of the stakeholders). Next, Borning et al. committed to three specific moral values to be supported explicitly. One is fairness, and more specifically freedom from bias. The simulation should not discriminate unfairly against any group of stakeholders, or privilege one mode of transportation or policy over another. A second is accountability. Insofar as possible, stakeholders should be able to confirm that their values are reflected in the simulation, evaluate and judge its validity, and develop an appropriate level of confidence in its output. The third is democracy. The simulation should support the democratic process in the context of land use, transportation, and environmental planning. In turn, as part of supporting the democratic process, Borning et al. decided that the model should not *a priori* favor or rule out any given set of stakeholder values, but instead, should allow different stakeholders to articulate the values that are most important to them, and evaluate the alternatives in light of these values.

4.3.2 Handling Widely Divergent and Potentially Conflicting Stakeholder Values. From the standpoint of conceptual investigations, UrbanSim as a design space poses tremendous challenges. The research team cannot focus on a few key values, as occurred in the Web Browser project (e.g., the value of informed consent), or the Room with a View project (e.g., the values of privacy in public spaces, and physical and psychological well-being). Rather, disputing stakeholders bring to the table widely divergent values about environmental, political, moral, and personal issues. Examples of stakeholder values are environmental sustainability, walkable neighborhoods, space for business expansion, affordable housing, freight mobility, minimal government intervention, minimal commute time, open space preservation, property rights, and environmental justice. How does one characterize the wide-ranging and deeply held values of diverse stakeholders, both present and future? Moreover, how does one prioritize the values implicated in the decisions? And how can one move from values to measurable outputs from the simulation to allow stakeholders to compare alternative scenarios?

As part of addressing these questions, the research group implemented a web-based interface that groups indicators into three broad value categories pertaining to the domain of urban development (economic, environmental, and social), and more specific value categories under that. To allow stakeholders to evaluate alternative urban futures, the interface provides a large collection of *indicators*: variables that distill some attribute of interest about the results [Gallopin 1997]. (Examples of indicators are the number of acres of rural land converted to urban use each year, the degree of poverty segregation, or the mode share between autos and transit.) These categories and indicators draw on a variety of sources, including empirical research on people's environmental concepts and values [Kahn 1999; Kahn and Kellert 2002], community-based indicator projects [Palmer 1998; Hart 1999], and the policy literature. Stakeholders can then use the interface to select indicators that speak to values that are important to them from among these categories.

This interface illustrates the interplay among conceptual, technical, and empirical investigations. The indicators are chosen to speak to different stakeholder values – responding to our distinction between explicitly supported values and stakeholder values in the initial conceptual investigation. The value categories are rooted empirically in both human psychology and policy studies, not just philosophy – and then embodied in a technical artifact (the web-based interface), which is in turn evaluated empirically.

4.3.3 Technical Choices Driven by Initial and Emergent Value Considerations. Most of the technical choices in the design of the UrbanSim software are in response to the need to generate indicators and other evaluation measures that respond to different strongly-held stakeholder values. For example, for some stakeholders, walkable, pedestrian-friendly neighborhoods are very important. But being able to model walking as a transportation mode makes difficult demands on the underlying simulation, requiring a finer-grained spatial scale than is needed for modeling automobile transportation alone. In turn, being able to answer questions about walking as a transportation mode is important for two explicitly supported values: fairness (not to privilege one transportation mode over another), and democracy (being able to answer questions about a value that is important to a significant number of stakeholders). As a second example of technical choices being driven by value considerations, UrbanSim's software architecture is designed to support rapid evolution in response to changed or additional requirements. For instance, the software architecture decouples the individual component models as much as possible, allowing them to evolve and new ones to be added in a modular fashion. Also, the system writes the simulation results into an SQL database, making it easy to write queries that produce new indicators quickly and as needed, rather than embedding the indicator computation code in the component models themselves. For similar reasons, the UrbanSim team uses the YP agile software development methodology [Freeman-Benson and Borning 2003], which allows the system to evolve and respond quickly to emerging stakeholder values and policy considerations.

4.3.4 Designing for Credibility, Openness, and Accountability. Credibility of the system is of great importance, particularly when the system is being used in a politically charged situation and is thus the subject of intense scrutiny. The research group has undertaken a variety of activities to help foster credibility, including using behaviorally transparent simulation techniques (i.e., simulating agents in the urban environment, such as households, businesses, and real estate developers, rather than using some more abstract and opaque simulation technique), and performing sensitivity analyses [Franklin et al. 2002] and a historical validation. In the historical validation, for example, the group started the model with 1980 data from Eugene/Springfield, simulated through 1994, and compared the simulation output with what actually happened. One of these comparisons is shown in Figure 3. In addition, our techniques for fostering openness and accountability are also intended to support credibility. These include using Open Source software (releasing the source code along with the executable), writing the code in as clear and understandable a fashion as possible, using a rigorous and extensive testing methodology, and complementing the Open Source software with an Open Process that makes the state of our development visible to anyone interested. For example, in our laboratory, a battery of tests is run whenever a new version of the software is committed to the source code repository. A traffic light (a real one) is activated by the testing regime – green means that the system has passed all tests, yellow means testing is under way, and red means that a test has failed. There is also a virtual traffic light, mirroring the physical one, visible on the web (www.urbansim.org/fireman). Similarly, the bug reports, feature requests, and plans are all on the UrbanSim project website as well. Details of this Open Process approach may be found in Freeman-Benson and Borning [2003].

Thus, in summary, Borning et al. are using Value Sensitive Design to investigate how a technology – an integrated land use, transportation, and environmental computer simulation – affects human values on both the individual and organizational levels; and how human values can continue to drive the technical investigations, including refining the simulation, data, and interaction model. Finally, employing Value Sensitive Design in a project of this scope serves to validate its use for complex, large-scale systems.

5. VALUE SENSITIVE DESIGN'S CONSTELLATION OF FEATURES

Value Sensitive Design shares and adopts many interests and techniques from related approaches to values and system design – computer ethics, social informatics, CSCW, and Participatory Design – as discussed in Section 2.2. However, Value Sensitive Design itself brings forward a unique constellation of eight features.

First, Value Sensitive Design seeks to be proactive: to influence the design of technology early in and throughout the design process.

Second, Value Sensitive Design enlarges the arena in which values arise to include not only the work place (as traditionally in the field of CSCW), but also education, the home, commerce, online communities, and public life.

Third, Value Sensitive Design contributes a unique methodology that employs conceptual, empirical, and technical investigations, applied iteratively and integratively (see Section 3).

Fourth, Value Sensitive Design enlarges the scope of human values beyond those of cooperation (CSCW) and participation and democracy (Participatory Design) to include all values, especially those with moral import. By moral, we refer to issues that pertain to fairness, justice, human welfare and virtue, encompassing within moral philosophical theory deontology [Dworkin 1978; Gewirth 1978; Kant 1785/1964; Rawls 1971], consequentialism ([Smart and Williams 1973]; see Scheffler [1982] for an analysis), and virtue [Foot 1978; MacIntyre 1984; Campbell and Christopher 1996]. Value Sensitive Design also accounts for conventions (e.g., standardization of protocols) and personal values (e.g., color preferences within a graphical user interface).

Fifth, Value Sensitive Design distinguishes between usability and human values with ethical import. Usability refers to characteristics of a system that make it work in a functional sense, including that it is easy to use, easy to learn, consistent, and recovers easily from errors [Adler and Winograd 1992; Norman 1988; Nielsen 1993]. However, not all highly usable systems support ethical values. Nielsen [1993], for example, asks us to imagine a computer system that checks for fraudulent applications of people who are applying for unemployment benefits by asking applicants numerous personal questions, and then checking for inconsistencies in their responses. Nielsen's point is that even if the system receives high usability scores some people may not find the system socially acceptable, based on the moral value of privacy.

Sixth, Value Sensitive Design identifies and takes seriously two classes of stakeholders: direct and indirect. Direct stakeholders refer to parties – individuals or organizations – who interact directly with the computer system or its output. Indirect stakeholders refer to all other parties who are affected by the use of the system. Often, indirect stakeholders are ignored in the design process. For example, computerized medical records systems have often been designed with many of the direct stakeholders in mind (e.g., insurance companies, hospitals, doctors, and nurses), but with too little regard for the values, such as the value of privacy, of a rather important group of indirect stakeholders: the patients.

Seventh, Value Sensitive Design is an interactional theory: values are viewed neither as inscribed into technology (an endogenous theory), nor as simply transmitted by social forces (an exogenous theory). Rather, the interactional position holds that while the features or properties that people design into technologies more readily support certain values and hinder others, the technology's actual use depends on the goals of the people interacting with it. A screwdriver, after all, is well-suited for turning screws, and is also amenable to use as a poker, pry bar, nail set, cutting device, and tool to dig up weeds, but functions poorly as a ladle, pillow, or wheel. Similarly, an online calendar system that displays individuals' scheduled events in detail readily supports accountability within an organization but makes privacy difficult. Moreover, through human interaction, technology itself changes over time. On occasion, such changes (as emphasized in the exogenous position) can mean the societal rejection of a technology, or that its acceptance is delayed. But more often it entails an iterative process whereby technologies are first invented, and then redesigned based on user interactions, which then are reintroduced to users, further interactions occur, and further redesigns implemented. Typical software updates (e.g., of word processors, browsers, and operating systems) epitomize this iterative process.

Eighth, Value Sensitive Design builds from the psychological proposition that certain values are universally held, although how such values play out in a particular culture at a particular point in time can vary considerably [Kahn 1999; Turiel 1998, 2002]. For example, even while living in an igloo, Inuits have conventions that ensure some forms of privacy; yet such forms of privacy are not maintained by separated rooms, as they are in most Western cultures. Generally, the more concretely (act-based) one conceptualizes a value, the more one will be led to recognizing cultural variation; conversely, the more abstractly one conceptualizes a value, the more one will be led to recognizing universals. Value Sensitive Design seeks to work both levels, the concrete and abstract, depending on the design problem at hand. Note that this is an empirical proposition, based on a large amount of psychological and anthropological data, not a philosophical one. We also make this claim only for certain values, not all – there are clearly some values that are culture-specific.

The three case studies presented in Section 5 illustrate the different features in this constellation. For example, UrbanSim illustrates the goal of being proactive and influencing the design of the technology early in and throughout the design process (Feature 1), and also involves enlarging the arena in which values arise to include urban planning and democratic participation in public decision-making (Feature 2). The cookies work is a good illustration of Value Sensitive Design's tripartite methodology (Feature 3): conceptual, technical, and empirical investigations, applied iteratively and integratively, were essential to the success of the project. Each of the three projects brings out a different set of human values (Feature 4): among others, informed consent for the cookies work; physical and psychological well-being and privacy in public spaces for Room with a View; and fairness, accountability, and democracy for UrbanSim, as well as the whole range of different sometimes competing stakeholder values. The cookies project illustrates the complex interaction between usability and human values (Feature 5): early versions of the system supported informed consent at the expense of usability, requiring additional work to develop a system that was both usable and provided reasonable support for informed consent. The Room with a View work considers and takes seriously both direct and indirect stakeholders (Feature 6): the occupants of the inside office ("The Watchers"), and passers-by in the plaza ("The Watched"). Value Sensitive Design's position that values are neither inscribed into technology nor simply transmitted by social forces (Feature 7) is illustrated by UrbanSim: the system by itself is certainly not neutral with respect to democratic process, but at the same time does not on its own ensure democratic decision-making on land use and transportation issues. Finally, the proposition that certain values are universally held, but play out in very different ways in different cultures and different times (Feature 8) is illustrated by the Room with a View project: the work is informed by a substantial body of work on the importance of privacy in all cultures (for example, the deep connection between privacy and self-identity), but concerns about privacy in public spaces play out in a specific way in the United States, and might do so quite differently in another cultural context.

We could draw out additional examples that illustrate Value Sensitive Design's constellation of features, both from the three case studies presented in Section 5, and in other projects; but hope that this short description demonstrates the unique contribution that Value Sensitive Design can make to the design of technology.

6. PRACTICAL SUGGESTIONS FOR USING VALUE SENSITIVE DESIGN

One natural question with Value Sensitive Design is, "How exactly do I do it?" In this section we offer some practical suggestions.

6.1. Start With a Value, Technology, or Context of Use

Any of these three core aspects – a value, technology, or context of use – easily motivates Value Sensitive Design. We suggest starting with the aspect that is most central to your work and interests. In the case of Informed Consent and Cookies, for example, Friedman et al. began with a value of central interest (informed consent) and moved from that value to its implications for Web browser design. In the case of UrbanSim, Borning et al. began with a technology (urban simulation) and a context of use (the urban planning process); upon inspection of those two, values issues quickly came to the fore.

6.2. Identify Direct and Indirect Stakeholders

As part of the initial conceptual investigation, systematically identify direct and indirect stakeholders. Recall that direct stakeholders are those individuals who interact directly with the technology or with the technology's output. Indirect stakeholders are those individuals who are also impacted by the system, though they never interact directly with it. In addition, it is worthwhile to recognize the following:

- Within each of these two overarching categories of stakeholders, there may be several subgroups.
- A single individual may be a member of more than one stakeholder group or subgroup. For example, in the UrbanSim project, an individual who works as an urban planner and lives in the area is both a direct stakeholder (i.e., through his or her direct use of the simulation to evaluate proposed transportation plans) and an indirect stakeholder (i.e., by virtue of living in the community for which the transportation plans will be implemented).
- An organizational power structure is often orthogonal to the distinction between direct and indirect stakeholders. For example, there might be low-level employees who are either direct or indirect stakeholders and who don't have control over using the system (e.g., workers on an assembly line). Participatory Design has contributed a substantial body of analysis to these issues, as well as techniques for dealing with them, such as ways of equalizing power among groups with unequal power. (See the references cited in Section 2.1.)

6.3. Identify Benefits and Harms for Each Stakeholder Group

Having identified the key stakeholders, systematically identify the benefits and harms for each group. In doing so, we suggest attention to the following points:

- Indirect stakeholders will be benefited or harmed to varying degrees; and in some designs it is probably possible to claim every human as an indirect stakeholder of some sort. Thus, one rule of thumb in the conceptual investigation is to give priority to indirect stakeholders who are strongly affected, or to large groups that are somewhat affected.
- Attend to issues of technical, cognitive, and physical competency. For example, children or the elderly might have limited cognitive competency. In such a case, care must be taken to ensure that their interests are represented in the design process, either by representatives from the affected groups themselves or, if this is not possible, by advocates.

- Personas [Pruitt and Grudin 2003] are a popular technique that can be useful for identifying the benefits and harms to each stakeholder group. However, we note two caveats. First, personas have a tendency to lead to stereotypes because they require a list of "socially coherent" attributes to be associated with the "imagined individual." Second, while in the literature each persona represents a different user group, in Value Sensitive Design (as noted above) the same individual may be a member of more than one stakeholder group. Thus, in our practice, we have deviated from the typical use of personas that maps a single persona onto a single user group, to allow for a single persona to map onto to multiple stakeholder groups.

6.4. Map Benefits and Harms onto Corresponding Values

With a list of benefits and harms in hand, one is in a strong position to recognize corresponding values. Sometimes the mapping is one of identity. For example, a harm that is characterized as invasion of privacy maps onto the value of privacy. Other times the mapping is less direct if not multifaceted. For example, with the Room with a View study, it is possible that a direct stakeholder's mood is improved when working in an office with an augmented window (as compared with no window). Such a benefit potentially implicates not only the value of psychological welfare, but also creativity, productivity, and physical welfare (health), assuming there is a causal link between improved mood and these other factors.

In some cases, the corresponding values will be obvious, but not always. Table 1 in Section 5.8 provides a table of human values with ethical import often implicated in system design. This table may be useful in suggesting values that should be considered in the investigation.

6.5. Conduct a Conceptual Investigation of Key Values

Following the identification of key values in play, a conceptual investigation of each can follow. Here it is helpful to turn to the relevant literature. In particular, the philosophical ontological literature can help provide criteria for what a value is, and thereby how to assess it empirically. (For example, Section 4.1.1 described how existing literature helped provide criteria for the value of informed consent.)

6.6. Identify Potential Value Conflicts

Values often come into conflict. Thus, once key values have been identified and carefully defined, a next step entails examining potential conflicts. For the purposes of design, value conflicts should usually not be conceived of as "either/or" situations, but as constraints on the design space. Admittedly, at times designs that support one value directly hinder support for another. In those instances, a good deal of discussion among the stakeholders may be warranted to identify the space of workable solutions. Typical value conflicts include accountability vs. privacy, trust vs. security, environmental sustainability vs. economic development, privacy vs. security, and hierarchical control vs. democratization.

6.7. Integrate Value Considerations Into One's Organizational Structure

Ideally, Value Sensitive Design will work in concert with organizational objectives. Within a company, for example, designers would bring values into the forefront, and in the process generate increased revenue, employee satisfaction, customer loyalty, and other desirable outcomes for their companies. In turn, within a government agency, designers would both better support national and community values, and enhance the organization's ability to achieve its objectives. In the real world, of course, human values (especially those with ethical import) may collide with economic objectives, power, and other factors. However, even in such situations, Value Sensitive Design should be able to make positive contributions, by showing alternate designs that better support enduring human values. For example, if a standards committee were considering adopting a protocol that raised serious privacy concerns, a Value Sensitive Design analysis and design might result in an alternate protocol that better addressed the issue of privacy while still retaining other needed properties. Citizens, advocacy groups, staff members, politicians, and others could then have a more effective argument against a claim that the proposed protocol was the only reasonable choice.

6.8. Human Values (with Ethical Import) Often Implicated in System Design

We stated earlier that while all values fall within its purview, Value Sensitive Design emphasizes values with ethical import. In Table 1, we present a list of frequently implicated values. This table is intended as a heuristic for suggesting values that should be considered in the investigation – it is definitely not intended as a complete list of human values that might be implicated.

Table 1. Human Values (with Ethical Import) Often Implicated in System Design

Human Value	Definition	Sample Literature
Human Welfare	Refers to people's physical, material, and psychological well-being	Leveson [1991]; Friedman, Kahn, & Hagman [2003]; Neumann [1995]; Turiel [1983, 1998]
Ownership and Property	Refers to a right to possess an object (or information), use it, manage it, derive income from it, and bequeath it	Becker [1977]; Friedman [1997b]; Herskovits [1952]; Lipinski & Britz [2000]
Privacy	Refers to a claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others	Agre and Rotenberg [1998]; Bellotti [1998]; Boyle, Edwards, & Greenberg [2000]; Friedman [1997b]; Fuchs [1999]; Jancke, Venolia, Grudin, Cadiz, and Gupta [2001]; Palen & Dourish [2003]; Nissenbaum [1998]; Phillips [1998]; Schoeman [1984]; Svensson, Hook, Laaksolahti, & Waern [2001]
Freedom From Bias	Refers to systematic unfairness perpetrated on individuals or groups, including pre-existing social bias, technical bias, and emergent social bias	Friedman & Nissenbaum [1996]; cf. Nass & Gong [2000]; Reeves & Nass [1996]
Universal Usability	Refers to making all people successful users of information technology	Aberg & Shahmehri [2001]; Shneiderman [1999, 2000]; Cooper & Rejmer [2001]; Jacko, Dixon, Rosa, Scott, & Pappas [1999]; Stephanidis [2001]
Trust	Refers to expectations that exist between people who can experience good will, extend good will toward others, feel vulnerable, and experience betrayal	Baier [1986]; Camp [2000]; Dieberger, Hook, Svensson, & Lonnqvist [2001]; Egger [2000]; Fogg & Tseng [1999]; Friedman, Kahn, & Howe [2000]; Kahn & Turiel [1988]; Mayer, Davis, &

		Schoorman [1995]; Olson & Olson [2000]; Nissenbaum [2001]; Rocco [1998]
Autonomy	Refers to people's ability to decide, plan, and act in ways that they believe will help them to achieve their goals	Friedman & Nissenbaum [1997]; Hill [1991]; Isaacs, Tang, & Morris [1996]; Suchman [1994]; Winograd [1994]
Informed Consent	Refers to garnering people's agreement, encompassing criteria of disclosure and comprehension (for "informed") and voluntariness, competence, and agreement (for "consent")	Faden & Beauchamp [1986]; Friedman, Millett, & Felten [2000]; The Belmont Report [1978]
Accountability	Refers to the properties that ensures that the actions of a person, people, or institution may be traced uniquely to the person, people, or institution	Friedman & Kahn [1992]; Friedman & Millet [1995]; Reeves & Nass [1996]
Courtesy	Refers to treating people with politeness and consideration	Bennett & Delatree [1978]; Wynne & Ryan [1993]
Identity	Refers to people's understanding of who they are over time, embracing both continuity and discontinuity over time	Bers, Gonzalo-Heydrich, & DeMaso [2001]; Rosenberg [1997]; Schiano & White [1998]; Turkle [1996]
Calmness	Refers to a peaceful and composed psychological state	Friedman & Kahn [2003]; Weiser & Brown [1997]
Environmental Sustainability	Refers to sustaining ecosystems such that they meet the needs of the present without compromising future generations	United Nations [1992]; World Commission on Environment and Development [1987]; Hart [1999]; Moldan, Billharz, & Matravers [1997]; Northwest Environment Watch [2002]

Two caveats. First, not all of these values are fundamentally distinct from one another. Nonetheless, each value has its own language and conceptualizations within its respective field, and thus warrants separate treatment here. Second, as noted above, this list is not comprehensive. Perhaps no list could be, at least within the confines of a paper. Peacefulness, respect, compassion, love, warmth, creativity, humor, originality, vision, friendship, cooperation, collaboration, purposefulness, devotion, loyalty, diplomacy, kindness, musicality, harmony – the list of other possible moral and non-moral values could get very long very quickly. Our particular list comprises many of the values that hinge on the deontological and consequentialist moral orientations noted above: human welfare, ownership and property, privacy, freedom from bias, universal usability, trust, autonomy, informed consent, and accountability. In addition, we have chosen several other values related to system design: courtesy, identity, calmness, and environmental sustainability.

6.9. Heuristics for Interviewing Stakeholders

As part of an empirical investigation, it is useful to interview stakeholders, to better understand their judgments about a context of use, an existing technology, or a proposed design. A semi-structured interview often offers a good balance between addressing the questions of interest and gathering new and unexpected insights. In these interviews, the following heuristics can prove useful:

- In probing stakeholders' reasons for their judgments, the simple question "Why?" can go a good distance. For example, seniors evaluating a ubiquitous computing video surveillance system might respond negatively to the system. When asked "Why?" a response might be: "I don't mind my family knowing that other people are visiting me, so they don't worry that I'm alone – I just don't want them to know who is visiting." The researcher can probe again: "Why don't you want them to know?" An answer might be: "I might have a new friend I don't want them to know about. It's not their business." Here the first "why" question elicits information about a value conflict (the family's desire to know about the senior's well-being and the senior's desire to control some information); the second "why" question elicits further information about the value of privacy for the senior.
- Ask about values not only directly, but indirectly, based on formal criteria specified in the conceptual investigation. For example, suppose that you want to conduct an empirical investigation of people's reasoning and values about "X" (say, trust, privacy, or informed consent), and that you decided to employ an interview methodology. One option is to ask people directly about the topic. "What is X?" "How do you reason about X?" "Can you give me an example from your own life of when you encountered a problem that involved X?" There is some merit to this direct approach. Certainly it gives people the opportunity to define the problem in their own terms. But you may quickly discover that it comes up short. Perhaps the greatest problem is that people have concepts about many aspects of the topic on which they cannot directly reflect. Rather, you will usually be better served by employing an alternative approach. As is common in social cognitive research (see Kahn [1999], chap. 5, for a discussion of methods), you could interview people about a hypothetical situation, or a common everyday event in their lives, or a task that you have asked them to solve, or a behavior in which they have just engaged. But, no matter what you choose, the important point is *a priori* to conceptualize what the topic entails, if possible demarcating its boundaries through formal criteria, and at a minimum employing issues or tasks that engage people's reasoning about the topic under investigation.

6.10. Heuristics for Technical Investigations

When engaging in value-oriented technical investigations, the following heuristics can prove useful:

- Technical mechanisms will often adjudicate multiple if not conflicting values, often in the form of design trade-offs. We have found it helpful to make explicit how a design trade-off maps onto a value conflict and differentially affects different groups of stakeholders. For example, the Room with a View study suggests real-time displays in interior offices may provide physiological benefits for those in the inside offices (the direct stakeholders), yet may impinge on the privacy and security of those walking through the outdoor scene (the indirect stakeholders), and especially women.
- Unanticipated values and value conflicts often emerge after a system is developed and deployed. Thus, when possible, design flexibility into the underlying technical architecture so that it can be responsive to such emergent concerns. In UrbanSim, for example, Borning et al. used agile programming techniques to design an architecture that can more readily accommodate new indicators and models.
- The control of information flow through underlying protocols – and the privacy concerns surrounding such control – is a strongly contested area. Ubiquitous

computing, with sensors that collect and then disseminate information at large, has only intensified these concerns. We suggest that underlying protocols that release information should be able to be turned off (and in such a way that the stakeholders are confident they have been turned off).

7. CONCLUSION

There is a growing interest and challenge to address values in design. Our goal in this paper has been to provide enough detail about Value Sensitive Design so that other researchers and designers can critically examine, use, and extend this approach. Our hope is that this approach can contribute to a principled and comprehensive consideration of values in the design of information and computational systems.

ACKNOWLEDGMENTS

Value Sensitive Design has emerged over the past decade and benefited from discussions with many people. We would like particularly to acknowledge all the members of our respective research groups, along with Edward Felten, Jonathan Grudin, Sara Kiesler, Clifford Nass, Helen Nissenbaum, John Thomas, and Terry Winograd. This research was supported in part by NSF Awards IIS-9911185, IIS-0325035, EIA-0121326, and EIA-0090832.

REFERENCES

ABERG, J., and SHAHMEHRI, N. 2001. An empirical study of human Web assistants: Implications for user support in Web information systems. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2000)* (pp. 404-411). New York, NY: Association for Computing Machinery Press.

ACKERMAN, M. S., and CRANOR, L. 1999. Privacy critics: UI components to safeguard users' privacy. In *Extended Abstracts of CHI 1999*, ACM Press, 258-259.

ADLER, P. S., and WINOGRAD, T., Eds. 1992. *Usability: Turning Technologies into Tools*. Oxford: Oxford University Press.

AGRE, P. E., and ROTENBERG, M., Eds. 1998. *Technology and Privacy: The New Landscape*. MIT Press, Cambridge, MA.

BAIER, A. 1986. Trust and antitrust. *Ethics*, 231-260.

BECK, A., and KATCHER, A. 1996. *Between Pets and People*. West Lafayette, IN: Purdue University Press.

BECKER, L. C. 1977. *Property Rights: Philosophical Foundations*. London, England: Routledge & Kegan Paul.

BELLOTTI, V. 1998. Design for privacy in multimedia computing and communications environments. In P. E. Agre and M. Rotenberg, Eds., *Technology and Privacy: The New Landscape* (pp. 63-98). Cambridge, MA: The MIT Press.

The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. 1978. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.

BENNETT, W. J., and DELATREE, E. J. 1978. Moral education in the schools. *The Public Interest*, *50*, 81-98.

BERS, M. U., GONZALEZ-HEYDRICH, J., and DEMASO, D. R. 2001. Identity construction environments: Supporting a virtual therapeutic community of pediatric patients undergoing dialysis. In *Proceedings of the Conference of Human Factors in Computing Systems (CHI 2001),* 380-387. New York, NY: Association for Computing Machinery.

BJERKNES, G., and BRATTETEIG, T. 1995. User participation and democracy: A discussion of Scandinavian research on system development. *Scandinavian Journal of Information Systems, 7*(1), 73-97.

BØDKER, S. 1990. *Through the Interface – A Human Activity Approach to User Interface Design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

BOYLE, M., EDWARDS, C., and GREENBERG, S. 2000. The effects of filtered video on awareness and privacy. In *Proceedings of Conference on Computer Supported Cooperative Work (CSCW 2000)*, 1-10. New York, NY: Association for Computing Machinery.

BYNUM, T. W., Ed. 1985. *Metaphilosophy, 16*(4). [Entire issue.]

CAMP, L. J. 2000. *Trust & Risk in Internet Commerce*. MIT Press, Cambridge, MA.

CAMPBELL, R. L., and CHRISTOPHER, J. C. 1996. Moral development theory: A critique of its Kantian presuppositions. *Developmental Review, 16*, 1-47.

COOPER, M., and REJMER, P. 2001. Case study: Localization of an accessibility evaluation. In *Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI 2001)*, 141-142. New York, NY: Association for Computing Machinery Press.

DIEBERGER, A., HOOK, K., SVENSSON, M., and LONNQVIST, P. 2001. Social navigation research agenda. In *Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI 2001)*, 107-108. New York, NY: Association of Computing Machinery Press.

DWORKIN, R. 1978. *Taking Rights Seriously*. Cambridge, MA: Harvard University Press.

EGGER, F. N. 2000. "Trust me, I'm an online vendor": Towards a model of trust for e-commerce system design. In *Extended Abstracts of the Conference of Human Factors in Computing Systems (CHI 2000),* 101-102. New York, NY: Association for Computing Machinery.

EHN, P. 1989. *Work-Oriented Design of Computer Artifacts*. Hillsdale, NJ: Lawrence Erlbaum Associates.

FADEN, R. and BEAUCHAMP, T. 1986. *A History and Theory of Informed Consent*. New York, NY: Oxford University Press.

FOGG, B.J., and TSENG, H. 1999. The elements of computer credibility. In *Proceedings of CHI 1999*, ACM Press, 80-87.

FOOT, P. 1978. *Virtues and Vices*. Berkeley and Los Angeles, CA: University of California Press.

FRANKENA, W. 1972. Value and valuation. In P. Edwards, Ed., The Encyclopedia of Philosophy, Vol. 7-8. (pp. 409-410). New York, NY: Macmillan.

FRANKLIN, J., WADDELL, P., and BRITTING, J. 2002. Sensitivity Analysis Approach for an Integrated Land Development & Travel Demand Modeling System, Presented at the Association of Collegiate Schools of Planning 44th Annual Conference, November 21-24, 2002, Baltimore, MD. Preprint available from www.urbansim.org.

FREEMAN-BENSON, B.N., and BORNING, A. 2003. YP and urban simulation: Applying an agile programming methodology in a politically tempestuous domain. In *Proceedings of the 2003 Agile Programming Conference*, Salt Lake City, June 2003. Preprint available from www.urbansim.org.

FRIEDMAN, B., Ed. 1997a. *Human Values and the Design of Computer Technology*. Cambridge University Press, New York NY.

FRIEDMAN, B. 1997b. Social judgments and technological innovation: Adolescents' understanding of property, privacy, and electronic information. *Computers in Human Behavior*, *13(*3), 327-351.

FRIEDMAN, B., HOWE, D. C., and FELTEN, E. 2002. Informed consent in the Mozilla browser: Implementing Value-Sensitive Design. In *Proceedings of HICSS-35,* IEEE Computer Society, Abstract, p. 247; CD-ROM of full papers, OSPE101.

FRIEDMAN, B. and KAHN, P. H., JR. 1992. Human agency and responsible computing: Implications for computer system design. *Journal of Systems Software, 17,* 7-14.

FRIEDMAN, B., KAHN, P. H., JR., and HOWE, D. C. 2000. Trust online. *Commun. ACM, 43,* 12, 34-40.

FRIEDMAN, B., and KAHN, P. H., JR. 2003. Human values, ethics, and design. In J. Jacko and A. Sears, Eds., *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Associates, Mahwah NJ.

FRIEDMAN, B., KAHN, P. H., JR., AND HAGMAN, J. 2003. Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship. *Conference Proceedings of CHI 2003*, 273 – 280. New York, NY: ACM Press.

FRIEDMAN, B., KAHN, P. H., JR., AND HAGMAN, J. 2004. The Watcher and The Watched: Social judgments about privacy in a public place. *Online Proceedings of CHI Fringe 2004*. Vienna, Austria: ACM CHI Place, 2004. (http://www.chiplace.org/chifringe/2004/198).

FRIEDMAN, B., and MILLETT, L. 1995. "It's the computer's fault" – Reasoning about computers as moral agents. In *Conference Companion of the Conference on Human Factors in Computing Systems (CHI 95)* (pp. 226-227). New York, NY: Association for Computing Machinery Press.

FRIEDMAN, B., MILLETT, L., and FELTEN, E. 2000. *Informed Consent Online: A Conceptual Model and Design Principles*. University of Washington Computer Science & Engineering Technical Report 00-12-2.

FRIEDMAN, B., and NISSENBAUM, H. Bias in computer systems. 1996. *ACM Transactions on Information Systems*, *14,* 3, 330-347.

FRIEDMAN, B., and NISSENBAUM, H. 1997. Software agents and user autonomy. *Proceedings of the First International Conference on Autonomous Agents*, 466-469. New York, NY: Association for Computing Machinery Press.

FUCHS, L. 1999. AREA: A cross-application notification service for groupware. In *Proceedings of ECSCW 1999*, Kluwer, Dordrechet Germany, 61-80.

GALEGHER, J., KRAUT, R. E., and EGIDO, C., Eds. 1990. *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*. Hillsdale, NJ: Lawrence Erlbaum Associates.

GALLOPIN, G.C. 1997. Indicators and their use: Information for decision-making. In B. Moldan, S. Billharz and R. Matravers, Eds., *Sustainability Indicators: A Report on the Project on Indicators of Sustainable Development*, Wiley, Chichester, England.

GEWIRTH, A. 1978. *Reason and Morality*. Chicago, IL: University of Chicago Press.

GREENBAUM, J., and KYNG, M., Eds. 1991. *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

GRUDIN, J. 1988. Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW '88),* 85-93. New York, NY: Association for Computing Machinery Press.

HAGMAN, J., HENDRICKSON, A., and WHITTY, A. 2003. What's in a barcode: Informed consent and machine scannable driver licenses. *CHI 2003 Extended Abstracts of the Conference on Human Factors in Computing System,* 912-913. New York, NY: ACM Press.

HART, M. 1999. *Guide to Sustainable Community Indicators*, Hart Environmental Data, PO Box 361, North Andover, MA 01845, second edition.

HERSKOVITS, M. J. 1952. *Economic Anthropology: A Study of Comparative Economics*. New York, NY: Alfred A. Knopf.

HILL, T. E., JR. 1991. *Autonomy and self-respect*. Cambridge: Cambridge University Press.

ISAACS, E. A., TANG, J. C., and MORRIS, T. 1996. Piazza: A desktop environment supporting impromptu and planned interactions. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW 96)*, 315-324. New York, NY: Association for Computing Machinery Press.

JACKO, J. A., DIXON, M. A., ROSA, R. H., JR., SCOTT, I. U., and PAPPAS, C. J. 1999. Visual profiles: A critical component of universal access. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 99)*, 330-337. New York, NY: Association for Computing Machinery Press.

JANCKE, G., VENOLIA, G. D., GRUDIN, J., CADIZ, J. J. and GUPTA, A. 2001. Linking public spaces: Technical and social issues. In *Proceedings of CHI 2001,* 530-537.

JOHNSON, E. H. 2000. Getting beyond the simple assumptions of organization impact [social informatics]. *Bulletin of the American Society for Information Science, 26*, 3, 18-19.

JOHNSON, D. G., and MILLER, K. 1997. Ethical issues for computer scientists and engineers. In A. B. Tucker, Jr., Ed.-in-Chief, *The Computer Science and Engineering Handbook* (pp. 16-26). CRC Press.

KAHN, P. H., JR. 1999. *The Human Relationship with Nature: Development and Culture*. MIT Press, Cambridge MA.

KAHN, P. H., JR., and KELLERT, S. R., Eds. 2002. *Children and Nature: Psychological, Sociocultural, and Evolutionary Investigations*. MIT Press, Cambridge MA.

KAHN, P. H., JR., and TURIEL, E. 1988. Children's conceptions of trust in the context of social expectations. *Merrill-Palmer Quarterly*, *34*, 403-419.

KANT, I. 1964. *Groundwork of the Metaphysic of Morals* (H. J. Paton, Trans.). New York, NY: Harper Torchbooks. (Original work published 1785.)

KLING, R., ROSENBAUM, H., and HERT, C. 1998. Social informatics in information science: An introduction. *Journal of the American Society for Information Science*, 49(12), 1047-1052.

KLING, R., and STAR, S. L. 1998. Human centered systems in the perspective of organizational and social informatics. *Computers and Society*, 28(1), 22-29.

KYNG, M., and MATHIASSEN, L., Eds. 1997. *Computers and Design in Context*. Cambridge, MA: The MIT Press.

LEVESON, N. G. 1991. Software safety in embedded computer systems. *Commun. ACM, 34*, 2, 34-46.

LIPINSKI, T. A., and BRITZ, J. J. 2000. Rethinking the ownership of information in the 21st century: Ethical implications. *Ethics and Information Technology, 2*, 1, 49-71.

MACINTYRE, A. 1984. *After Virtue*. Nortre Dame: University of Nortre Dame Press.

MAYER, R.C., DAVIS, J.H, AND SCHOORMAN, F.D., 1995. An integrative model of organizational trust. *The Academy of Management Review, 20*, 3, 709-734.

MILLETT, L., FRIEDMAN, B., and FELTEN, E. 2001. Cookies and web browser design: Toward realizing informed consent online. In *Proceedings of CHI 2001,* ACM Press, 46-52.

MOLDAN, B., BILLHARZ, S., and MATRAVERS, R., Eds., 1997. *Sustainability Indicators: A Report on the Project on Indicators of Sustainable Development*, Wiley, Chichester, England.

MOORE, G. E. 1978. *Principia ethica*. Cambridge: Cambridge University Press. (Original work published 1903.)

NASS, C., and GONG, L. 2000. Speech interfaces from an evolutionary perspective. *Communications of the ACM, 43*(9), 36-43.

NEUMANN, P. G. 1995. *Computer Related Risks*. New York, NY: Association for Computing Machinery Press.

NIELSEN, J. 1993. *Usability Engineering*. Boston, MA: AP Professional.

NISSENBAUM, H. 1998. Protecting privacy in an information age: The problem with privacy in public. *Law and Philosophy, 17*, 559-596.

NISSENBAUM, H. 1999. Can trust be secured online? A theoretical perspective. *Etica e Politca*, *2* (Electronic journal).

NISSENBAUM, H. 2001. Securing trust online: Wisdom or oxymoron. *Boston University Law Review, 81(3), 635-664*.

NORMAN, D. A. 1988. *The Psychology of Everyday Things*. New York, NY: Basic Books.

NORTHWEST ENVIRONMENT WATCH. 2002. *This Place on Earth 2002: Measuring What Matters*. Northwest Environment Watch, 1402 Third Avenue, Seattle, WA 98101.

NOTH, M., BORNING, A., and WADDELL, P. 2003. An extensible, modular architecture for simulating urban development, transportation, and environmental impacts. *Computers, Environment and Urban Systems*, *27*, 2, 181-203.

OLSON, J. S., and OLSON, G. M. 2000. i2i trust in e-commerce. *Communications of the ACM, 43*(12), 41-44.

OLSON, J. S., and TEASLEY, S. 1996. Groupware in the wild: Lessons learned from a year of virtual collaboration. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW 96)*, 419-427. New York, NY: Association for Computing Machinery Press.

ORLIKOWSI, W. J., and IACONO, C. S. 2001. Research commentary: desperately seeking the "IT" in IT research—a call to theorizing the IT artifact. *Information Systems Research, 12*, 2, 121-134.

PALEN, L., and GRUDIN, J. 2003. Discretionary adoption of group support software: Lessons from calendar applications. In B.E. Munkvold, Ed., *Implementing Collaboration Technologies in Industry,* Springer Verlag, Heidelberg.

PALEN, L., and DOURISH, P. 2003. Privacy and trust: Unpacking "privacy" for a networked world. In *Proceedings of CHI 2003,* 129-136.

PALMER, K., Ed. 1998. *Indicators of Sustainable Community*, Sustainable Seattle, Seattle, WA.

PHILLIPS, D. J. 1998. Cryptography, secrets, and structuring of trust. In P. E. Agre and M. Rotenberg, Eds., *Technology and Privacy: The New Landscape* (pp. 243-276). Cambridge, MA: The MIT Press.

PRUITT, J., and GRUDIN, J. 2003. Personas: Practice and theory. In *Proceedings of DUX 2003*, ACM Press.

RAWLS, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

REEVES, B., and NASS, C. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY and Stanford, CA: Cambridge University Press and CSLI Publications.

RIEGELSBERGER, J., and SASSE, M. A. 2002. Face it – Photos don't make a web site trustworthy, in *Extended Abstracts of CHI 2002,* ACM Press, 742-743.

ROCCO, E. 1998. Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *Proceedings of CHI 1998*, ACM Press, 496-502.

ROSENBERG, S. 1997. Multiplicity of selves. In R. D. Ashmore and L. Jussim, Eds., *Self and Identity: Fundamental Issues* (pp. 23-45). New York, NY: Oxford University Press.

SAWYER, S, and ROSENBAUM, H. 2000. Social informatics in the information sciences: Current activities and emerging direction. *Informing Science, 3*(2), 89-95.

SCHEFFLER, S. 1982. *The Rejection of Consequentialism*. Oxford, England: Oxford University Press.

SCHIANO, D. J., and WHITE, S. 1998. The first noble truth of cyberspace: People are people (even when they MOO). In *Proceedings of the Conference of Human Factors in Computing Systems (CHI 98),* 352-359. New York, NY: Association for Computing Machinery.

SCHNEIDER, F. B., Ed. 1999. *Trust in Cyberspace*. National Academy Press, Washington, D.C.

SCHOEMAN, F. D., Ed. 1984. *Philosophical Dimensions of Privacy: An Anthology*. Cambridge, England: Cambridge University Press.

SHNEIDERMAN, B. 1999. Universal usability: Pushing human-computer interaction research to empower every citizen. ISR Technical Report 99-72. University of Maryland, Institute for Systems Research. College Park, MD.

SHNEIDERMAN, B. 2000. Universal usability. *Commun. of the ACM, 43*, 5, 84-91.

SIMPSON, J.A., and WEINER, E.S.C., Eds. 1989. "value, *n*." *Oxford English Dictionary*. Oxford: Clarendon Press, 1989. *OED Online*. Oxford University Press. 30 May 2003. http://dictionary.oed.com/cgi/entry/00274678

SMART, J. J. C. and WILLIAMS, B. 1973. *Utilitarianism For and Against*. Cambridge: Cambridge University Press.

STEPHANIDIS, C., Ed. 2001. *User Interfaces for All: Concepts, Methods, and Tools*. Mahwah, NJ: Lawrence Erlbaum Associates.

SUCHMAN, L. 1994. Do categories have politics? The language/action perspective reconsidered. *CSCW Journal, 2*, 3 , 177-190.

SVENSSON, M., HOOK, K., LAAKSOLAHTI, J., and WAERN, A. 2001. Social navigation of food recipes. In *Proceedings of the Conference of Human Factors in Computing Systems (CHI 2001),* 341-348. New York, NY: Association for Computing Machinery.

TANG, J. C. 1997. Eliminating a hardware switch: Weighing economics and values in a design decision. In B. Friedman, Ed., *Human Values and the Design of Computer Technology,* (pp. 259-269). Cambridge Univ. Press, New York NY.

THOMAS, J. C. 1997. Steps toward universal access within a communications company. In B. Friedman, Ed., *Human Values and the Design of Computer Technology,* (pp. 271-287). Cambridge Univ. Press, New York NY.

TURIEL, E. 1983. *The Development of Social Knowledge*. Cambridge, England: Cambridge University Press.

TURIEL, E. 1998. Moral development. In N. Eisenberg, Ed., *Social, Emotional, and Personality Development* (pp. 863-932). Vol. 3 of W. Damon, Ed., *Handbook of Child Psychology*. 5th edition. New York, NY: Wiley.

TURIEL, E. 2002. *The Culture of Morality: Social Development, Context, and Conflict*. Cambridge, England: Cambridge University Press.

TURKLE, S. 1996. *Life on the Screen: Identify in the Age of the Internet*. New York, NY: Simon and Schuster.

ULRICH, R. S. 1984. View through a window may influence recovery from surgery. *Science, 224*, 420-421.

ULRICH, R. S. 1993. Biophilia, biophobia, and natural landscapes. In S. R. Kellert and E. O. Wilson, Eds., *The Biophilia Hypothesis* (pp. 73-137). Washington, D.C.: Island Press.

UNITED NATIONS. 2002. Report of the United Nations Conference on Environment and Development, held in Rio de Janeiro, Brazil, 1992. Available from http://www.un.org/esa/sustdev/documents/agenda21/english/agenda21toc.htm

WADDELL, P. 2002. UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association*, *68*, 3, 297-314.

WADDELL, P., BORNING, A., NOTH, M., FREIER, N., BECKE, M., and ULFARSSON, G. 2003. Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics*, *3*, 1, 43-67.

WEISER, M., and BROWN, J. S. 1997. The coming age of calm technology. In P. Denning and B. Metcalfe, Eds., *Beyond Calculation: The Next 50 Years of Computing* (pp. 75-85). New York, NY: Springer-Verlag.

WEIZENBAUM, J. 1972. On the impact of the computer on society: How does one insult a machine? *Science, 178*, 609-614.

WIENER, N. 1985. The machine as threat and promise. In P. Masani, Ed., *Norbert Wiener: Collected Works and Commentaries, Vol. IV* (pp. 673-678). Cambridge, MA: The MIT Press. (Reprinted from *St. Louis Post Dispatch*, 1953, December 13.)

WINOGRAD, T. 1994. Categories, disciplines, and social coordination. *CSCW Journal, 2*, 3, 191-197.

WORLD COMMISSION ON ENVIRONMENT AND DEVELOPMENT (Gro Harlem Brundtland, Chair). 1987. *Our Common Future*. Oxford University Press, Oxford.

WYNNE, E. A., and RYAN, K. 1993. *Reclaiming our schools: A handbook on teaching character, academics, and discipline*. New York, Macmillan.

ZHENG, J., BOS, N., OLSON, J., and OLSON, G. M. 2001. Trust without touch: Jump-start trust with social chat. In *Extended Abstracts of CHI* 2001, ACM Press, 293-294.