Design Methods for Ethical Persuasive Computing

Janet Davis
Dept. of Computer Science
Grinnell College
Grinnell, IA 50112
davisjan@cs.grinnell.edu

ABSTRACT

Value Sensitive Design and Participatory Design are two methodological frameworks that account for ethical issues throughout the process of technology design. Through analysis and case studies, this paper argues that such methods should be applied to persuasive technology—computer systems that are intended to change behaviors and attitudes.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*ethics*; K.4.2 [**Computers and Society**]: Social Issues

General Terms

Design, Human Factors

Keywords

Persuasive technology, design methods, Value Sensitive Design, Participatory Design

1. INTRODUCTION

Though each of us encounters or engages in persuasive communication on an almost daily basis, some find the very idea of a persuasive computer alarming. Persuasive computing, or *captology*, is the study of "computers as persuasive technology," that is, the design of computer systems to change behaviors and attitudes [12, 13]. There is good reason to be concerned about information technology that is intended to change people's behavior. Yet, technologies *do* change how people live, and much of the persuasive technology designed thus far serves ends that most would agree are good for everyone: for example, to conserve energy [32], to be more physically active [7, 21], to stop smoking [19].

Even when designers have good intent, persuasive computing systems are at least as fraught with potential ethical missteps as other types of technologies. For example, the Breakaway desk sculpture, designed to gently remind information workers to take stretch breaks [21], is appealing in intent and charming in form. Yet, it may inadvertently cause its user harm to his or her reputation as a hard worker, because it reveals to any observer how recently the user left his or her chair. While such potential harms need not necessarily prevent the deployment of such technology, they ought to be considered.

I am not saying that ethical concerns have gone unrecognized in the persuasive computing community. Indeed, B.J. Fogg devotes an entire chapter to ethics in his 2003 book, *Persuasive Technology* [13]. In this chapter, Fogg cautions against ethical "red flags" in designing persuasive technology and sketches an approach to analyzing harms and benefits. Even earlier, in the *Communications of the ACM*'s May 1999 issue on persuasive technologies, Berdichevsky and Neuenschwander developed a set of ethical principles or guidelines for persuasive technology [2].

At the same time, persuasive computing systems are not the only ones that raise ethical issues. I would argue that principles and guidelines are not enough, and that persuasive technology designers should look to the broader human-computer interaction community for methods that help designers uncover and address ethical issues. In this paper, I consider two such methodological frameworks: Value Sensitive Design and Participatory Design. Value Sensitive Design guides designers to systematically address human values, such as privacy and autonomy, throughout the design process [17]; Participatory Design aims to engage key stakeholders as full and equal particicipants throughout the design process [11]. Both methods have potential to overcome ethical problems with persuasive technology.

The paper proceeds as follows. In the next two sections, I make the case for design methodology to address ethical issues in persuasive computing, first arguing that such issues are not as unique as they might appear, and second that design methodology can go beyond the application of principles alone. In turn, I explain the Value Sensitive Design and Participatory Design frameworks, considering the strengths of each in the context of persuasive technology. To support this analysis, I discuss relevant case students. Finally, I conclude with challenges and future work for applying these two frameworks to designing persuasive technology.

2. ETHICAL ISSUES IN PERSUASIVE COMPUTING: UNIQUE OR COMMON?

Attempting to change others' behaviors or attitudes seems an ethical minefield. When we are recipients of persuasive communications—imagine the salesperson, political activist, or solicitor for charity at your front door—we may not be-

PERSUASIVE '09, April 26-29, Claremont, California, USA

lieve that the persuader has our best interests in mind. Yet, we may feel under duress or social pressure to take a particular action (and we may resent this pressure). We may feel our emotions are tugged at unfairly, and we may not trust that we are being fully informed or told the truth. These concerns only increase when the target of persuasion is emotionally vulnerable—say, lonely or bereaved—or cognitively vulnerable—a child or an elderly person with dementia.

Yet, in my view and that of many others, including B.J. Fogg [13], persuasion is not inherently unethical and can serve social goods. Indeed, it is difficult for designers to avoid changing others' behaviors. Most, if not all, information technologies instigate complicated causal chains that do change people's behavior, even if the term "persuasion" never passes through the designers' mind [1]. Arguably, computers are just one medium among many for conveying persuasive communications.

That said, Fogg identifies six ethical issues that are unique to computers as a medium for persuasion [13, p. 213–220]. First, the novelty of a computer system can blind users to designers' persuasive intent. Second, designers can exploit computers' reputation as "intelligent and fair." Third, computers can be far more ubiquitous and persistent than a human persuader. Fourth, computers cannot be negotiated with; they do only what they are programmed to do. Fifth, computers can affect emotions but do not have emotions themselves. And finally, computers do not share in moral responsibility for harmful outcomes.

In response, I observe that while these ethical concerns are very real, they are not unique to *persuasive* computing systems. Similar issues are of deep concern for other types of information systems as well. For example, Friedman and Nissenbaum analyze the problem of bias, or the systematic and unfair assignment of undesirable outcomes to a person or group, in a range of information systems concerned with tasks from making airline reservations to determining whether persons are eligible for citizenship [18]. Computers' reputation as "intelligent and fair" is certainly a factor in how computer system bias is perceived and enacted, or whether it is noticed at all. Millet, Friedman, and colleagues consider informed consent and what it means in the context of distracted, goal-driven web surfers and infinitely persistent computers [28, 16]. Finally, Johnson and Mulvey examine moral and legal accountability in the use of computer decision-support systems [22]. Although they also conclude that computers cannot have moral responsibility, computers can obscure who *does* have that responsibility.

These examples show there are lessons to be learned from how philosophers and designers have analyzed and accounted for such ethical issues in information systems beyond persuasive technology. But, I have not yet argued for ethical design methods, as opposed to ethical principles.

3. ETHICAL PRINCIPLES AND METHODS

When faced with ethical dilemmas particular to a domain of technology development, a common approach is to develop ethical guidelines or principles. For example, consider the "covenants" for simulation systems: The covenants with reality and values specify that models must adequately represent the real world and that objective functions should optimize in accordance with the client's values [25], while the covenant with transparency requires that simulation models be open to scrutiny [10]. Furthermore, Ören et al. proposed a code of professional ethics for simulationists [31].

Indeed, in her critical review of Fogg's *Persuasive Technology* [13], Atkinson [1] argues that not enough attention is paid to ethical principles that encapsulate society's collective understanding of ethical problems. Atkinson pinpoints one key principle: that, for persuasion to be ethical, the audience must be informed of the persuaders' intent.

Disclosure is also identified as a key principle by Berdichevsky and Neuenschwander in their article, "Toward an Ethics of Persuasive Technology" [2]. Similar to Mason's "covenant with reality" [25], and prior to Fogg's 2003 text [13], Berdichevsky and Neuenschwander argued that persuasive technology must not misinform. Two further principles focus on designer's persuasive intent: designers' motivations must be such that they would be deemed ethical if no technology were involved, and, in their "Golden Rule," creators of technology should not seek to persuade others of something they would not be persuaded of themselves. Two principles concern privacy implications of persuasive technology: that creators of persuasive technology should protect users' privacy as their own, and that technologies relaying information to third parties should be subject to particular scrutiny. Finally, two principles pertain to outcomes: First, intended outcomes must be such that they would be ethical even without persuasion, and second, the creators of persuasive technology must anticipate and assume responsibility for all "reasonably predictable" outcomes.

But how can designers predict and anticipate outcomes for all stakeholders based on their own limited perspectives? How can privacy be protected (and what do we mean by privacy, anyway)? How do we ensure that individuals are informed of persuasive intent and consent to being persuaded?

While principles are certainly helpful, I would claim that they are often not enough. Designers need methods to structure their efforts to enact ethical principles. Moreover, methods can help designers discover what to do in cases where established principles are inadequate. Consider an analogy with usability: Although principles help designers to critique systems and to develop more usable designs in the first place, many problems are revealed only by watching what happens when users interact with the system. From many such observations, new principles emerge. As case in point, under the rubric of Value Sensitive Design, Millet, Friedman, and Felten used the legal and philosophical literature to develop five key requirements for technology to support the informed consent [28]; studying users revealed a sixth requirement, that of minimal distraction [16].

In support of methods for ethical design, Atkinson not only argued for an overarching ethical principle for persuasive technology; she also pointed out that designers can reduce "unintended and unforeseen consequences" through "public consultation," "social learning," and "multistakeholder negotiation" (B. Campbell, cited in [1]). Atkinson also cites scenarios and personas as methods that can help designers to envision technology in use and thus anticipate unintended consequences. However, as Johnson wrote in 2004, Fogg's 2003 book [13] is "designer centered and system centered," projecting a "view of the designer as all knowing" [23]; there is too little attention to methods that engage potential users and other stakeholders. While more

recent work in persuasive technology takes a user-centered approach (for example, [7, 19, 34]), I would argue that it does so from perspectives of usability and persuasive effectiveness, and not from an explicitly ethical standpoint. Further, persuasive technology can affect additional stakeholders beyond just the users. Fortunately, since the ethical problems of persuasive technology are not unique, the designer can draw upon established methods for information technology design that explicitly engage with ethical issues. For example, the empirical investigations of Value Sensitive Design help the designer to understand, not the usability of the technology, but stakeholders' beliefs about related human values, such as privacy or autonomy, and their moral reactions to the designed technology. Participatory Design actively engages future technology users in the design process, and is very much concerned with social learning and multi-stakeholder negotiation.

In the remainder of this paper, I will discuss the applicability of Value Sensitive Design and Participatory Design methods to the design of persuasive computing systems. I will present and explore each method in turn, considering relevant case studies from the literature.

4. VALUE SENSITIVE DESIGN

Value Sensitive Design (VSD), developed by Friedman and colleagues [14, 17], is a theoretical and methodological framework that seeks to account for human values in a principled and comprehensive way throughout the design process. VSD emphasizes values of moral import—values such as fairness, autonomy, privacy, and human welfare—and thus speaks to ethical concerns in technology design. Over the past fifteen years, researchers have applied VSD to such diverse values and technologies as informed consent in delivering web browser cookies [28, 16], democracy and fairness in an urban simulation system [8, 15], and reputation and trust in a corporate knowledge-sharing system [26].

Key features of Value Sensitive Design include its interactional perspective, attention to both direct and indirect stakeholders, and a tripartite methodology [17]. First, VSD is an interactional theory: people and social systems affect technological development, and technologies shape but do not determine individual behavior and social systems. Second, VSD requires attention to both direct and indirect stakeholders: not only those who use the technology, but those who are affected by its use. Third, the VSD framework employs a tripartite methodology, incorporating conceptual, technical, and empirical investigations in an iterative and integrative process. Conceptual investigations encompass theoretical and literature-based explorations of the stakeholders and values at hand; these are considered in greater detail in the next subsection. Technical investigations focus on the technology itself through careful analysis of how system features support or undermine particular values. Finally, empirical investigations consider stakeholders' understandings of values and their implications, along with the human response to the artifact; empirical investigations are significant to the case study discussed later in this section.

Value Sensitive Design has the potential to contribute to design of ethical persuasive computing in several significant ways. First, VSD's interactional perspective is important to persuasive technology: persuasive intent emerges from a social context, and while persuasive technologies are intended to change behavior, they cannot force the desired change. Second, VSD provides a more thorough method of stakeholder and value analysis. By shifting attention from potential harms to human values, VSD drills down to *why* a design might be considered harmful or beneficial, at the same time broadening the range of concerns and helping to reveal value conflicts or tensions that must be adjudicated. Third, VSD provides a growing body of methods for exploring and anticipating the value implications of a technology. Finally, by engaging with human values throughout the design process, VSD helps uncover and account for problems that might not otherwise be seen until the system is deployed. The remainder of this section considers these potential contributions in greater depth: first comparing VSD's stakeholder and value analysis to that presented by Fogg, then discussing the value scenario method, and finally considering knowledge sharing systems as a design case study.

4.1 Stakeholder and Value Analysis

Both Fogg [13, p. 233–235] and Friedman, Kahn, and Borning [17] recommend the use of stakeholder analysis to uncover a new system's ethical or value implications. While the methods presented are very similar, VSD's approach to conceptual investigations draws more attention to the welfare of indirect stakeholders, provides additional guidance in identifying values at stake, and reveals situations in which designers must make tradeoffs between conflicting value concerns. Furthermore, while Fogg sketches a method for stakeholder analysis, Value Sensitive Design literature provides a principled, systematic approach, along with several case studies of its application.

Both methods begin by identifying stakeholders. Direct stakeholders are the most obvious: those who interact directly with the technology. Fogg suggests that we also consider technology distributors and "sometimes those who are close to the users as well" [13, p. 233]. VSD takes indirect stakeholders—"those individuals who are also impacted by the system, though they never interact directly with it" [17, p. 362]—far more seriously. Indeed, indirect stakeholders, such as patients in the case of a medical technology, can be affected much more strongly than the direct stakeholders who actually control the technology. Technologies such as urban planning systems can reach out to very large and diffuse, but important, groups of indirect stakeholders, sometimes reaching out to all members of a society. Thus, no VSD stakeholder analysis is complete without a thorough consideration of indirect stakeholders, even if they are eventually considered to have a relatively small stake in the design. Friedman et al. also ask us to recall that an individual can have multiple stakeholder roles, for example as a planner and a resident, and that organizational power structures often crosscut direct and indirect stakeholder roles.

Both Fogg and Friedman et al. direct designers to systematically identify benefits and harms for each stakeholder. However, where Fogg suggests that the designer identify which stakeholder has the most to gain and which the most to lose, VSD distributes attention among many stakeholders. In particular, Friedman et al. suggest that designers give priority to all direct stakeholders, to the indirect stakeholders who are most strongly affected, and also to large groups of indirect stakeholders who are moderately affected. Both agree that special attention should be given to vulnerable groups, such as children and the elderly.

Next, both methods direct designers to identify values im-

plicated by the benefits and harms. However, Fogg provides little guidance for doing this. By contrast, in VSD the stakeholder analysis is a prelude to a much deeper examination of implicated values through the tripartite methodology. To help translate from harms and benefits to values, and to prompt the designer to consider additional values, Friedman et al. provide a table of key values often implicated in technology design [17, p. 364–365]. VSD's conceptual investigation does not end here; designers should go on to investigate key values through philosophical, psychological, legal, or other literature. Such literature can help to define a value, and thus provide key criteria for assessing the value in technical and empirical investigations. To help kickstart these investigations, Friedman et al. also provide a sampling of works that are concerned with each key value, drawing on a variety of disciplinary literatures.

For example, Fogg states that coercion and deception are categorically *not* persuasive strategies [13, p. 15], but why are they ethically problematic? I believe we can trace these concerns to a respect for human autonomy: a person's right to decide for him or herself how best to act [17]. Moreover, coercion raises questions regarding moral accountability for the computer's actions, since, as Fogg points out, the computer is not a moral agent able to make decisions in context; deception raises questions not only for an abstract notion of truth but also with respect to whether the user was provided an opportunity for informed consent.

This admittedly very brief analysis sheds some light on Fogg's "methods that raise red flags": operant conditioning and surveillance [13, p. 224]. Surveillance obviously has implications for privacy, as Fogg points out [13, p. 226]. Nonetheless, Fogg says, such technology could be ethical if it were set up mainly to be supportive or helpful; it is unethical if it is intended mainly to punish. Operant conditioning, the use of rewards and punishments to promote certain behaviors, is a powerful psychological phenomenon that may go on unnoticed, so powerful that it may be seen as an infringement on the user's autonomy. Again, Fogg is particularly concerned when the system incorporates punishment; I would trace this concern not only to respect for human autonomy but also to questions regarding moral accountability for the automated punishments. Fogg is less concerned when the user is informed [13, p. 225], but the designer should ask, how do we ensure the user is informed and consents to the treatment? Through careful analysis and empirical study under the rubric of Value Sensitive Design, Millet, Friedman, and colleagues developed requirements for informed consent online [28, 16], which can and should be applied here. Beyond informed consent, autonomy, accountability, and privacy, persuasive techniques may take advantage of or infringe upon values such as courtesy, trust, and identity.

Furthermore, VSD goes beyond Fogg's stakeholder analysis to directly consider the implications of the values. Value conflicts or tensions may constrain the design space; for example, a need for accountability may conflict with respect for privacy. And, drawing on the interactional perspective, VSD directs designers to support key values in the policies and practices surrounding the technology. Such considerations can be critical to the success of a new system, as in the case study later in this section.

Finally, both methods remind us to consider the designer's own values and assumptions. Both Fogg and Friedman et al. caution that the concrete expression and relative importance of values vary across cultures; our own assumptions may not apply in designing for other cultures. Beyond this, designers should consider the approach taken in Value Sensitive Design of UrbanSim, a large-scale urban planning system: to explicitly choose a set of values that the system will support, and take a principled approach to other stakeholder values [17]. Explicitly supported values are those for which the designer can make a principled argument that the value is universal or morally right for the system to support within the societal context of use. Designers taking such a principled approach commit to not privileging their own personal values, beyond the explicitly supported values, over the values of other stakeholders.

But, this analysis leaves us with a question: How can designers predict potential harms and benefits to a range of stakeholders, and thus discover the values at stake? The following subsection suggests one appropriate method from the Value Sensitive Design literature.

4.2 Method: Value Scenarios

As Atkinson argues, scenarios can be a powerful approach to predicting unintended effects of new technology [1]. Both Berdichevsky and Neuenschwander [2] and Nathan et al. [30] propose envisioning the negative implications of technologies from a design noir or "dark side" perspective. But where Berdichevsky and Neuenschwander's "dark side designs" provide a simple approach to help students and designers consider the darker side of persuasive technology in general, Nathan et al.'s "value scenarios" aim to help the designer anticipate systemic effects of a specific technology.

Both approaches are concerned with implications of the technology for ethical values, and both differ from traditional design scenarios in that they focus on negative rather than positive uses of the technology. However, Berdichevsky and Neuenschwander "dark side designs" focus on describing the technologies themselves. In the examples given, even the designer's intentions are, if not dark, at least dubious: to convert people to a new religion or to persuade children to divulge their secrets [2]. These technologies have disturbing ethical implications even when used exactly as designed.

By contrast, Nathan et al.'s value scenarios method involves writing stories about how people use a technology, often not as the designer intended [30]. While the technologies Nathan et al. explore arise from good intentions—for example, the hypothetical SafetyNet system is intended to help vulnerable people avoid unsafe neighborhoods—value scenarios focus on possible negative effects of those seemingly positive technologies. Nathan et al. accomplish this by considering three dimensions beyond the immediate implications for moral values: effects for both direct and indirect stakeholders; effects over long periods of time during which systemic effects can develop; and the pervasive effects of technology use that moves beyond an isolated user or group of users to cross space, cultures, and demographic groups. Nathan et al. use these rich and provocative dimensions to envision a very plausible world in which SafetyNet helps to racially segregate neighborhoods and enable hate crimes.

Thus, the VSD methodology provides a powerful tool for exploring unintended ethical implications of technology. VSD provides one answer to the question that, to my mind, is left unasked and unanswered by Berdichevsky and Neuenschwander [2] and by Fogg [13]: *How* can the designer predict unintended outcomes of persuasive technology?

4.3 Case Study: Knowledge Sharing

Finally, I consider a case study that demonstrates VSD's effectiveness in a persuasive computing context. Using the framework of Value Sensitive Design, Miller and colleagues developed the CodeCOOP groupware system to promote knowledge sharing among engineers at a large software company [26]. Although the system falls solidly within the area of groupware, it was designed with persuasive intent: to persuade engineers to contribute and seek out code and information across organizational boundaries, and ultimately to change organizational behaviors and attitudes around knowledge sharing. Indeed, Brodie et al. presented their work on a similar problem—promoting knowledge sharing among customer support workers and system administrators within an organization—at the Second International Conference on Persuasive Technology in 2007 [5].

Although both research teams attended to issues of reputation in their design process, the designers of CodeCOOP examined a richer field of values early on and contributed a method for adjudicating tensions revealed in empirical study of such values. Brodie et al. [5] analyze the knowledge sharing problem as a prisoner's dilemma: although everyone benefits when knowledge is shared, a self-interested employee may prefer to keep his own knowledge as a private asset while reaping the benefit of knowledge shared by others. If many or all employees act out of self-interest, then little sharing occurs. Brodie et al. propose reputation as a persuasive mechanism to increase the value of contributing code, so that sharing is worthwhile even to self-interested employees. By contrast, using VSD's stakeholder and value analysis method, Miller et al. identified a richer field of values including not only reputation but privacy, trust, and awareness [26]. Moreover, Miller et al. conceive of reputation as both a potential benefit and a potential harm: While contributing good information could benefit a user's reputation, posting incorrect information or asking simple questions could harm a user's reputation.

We also see differences in the empirical approach to human values and ethical concerns. Early empirical work at IBM focused on the potential usefulness of a knowledge-sharing system [33]. Only after deployment, when evaluating factors that might prevent users from reaching out to an expert, did concerns about anonymity and potential harms to reputation emerge. By contrast, empirical work by Miller et al. focused early on potential harms and benefits, and their relation to ethical values [26]. Careful analysis of the values allowed the designers to take a broader, more principled view of the potential role of each value; they asked several questions pertaining to each value, looking at both benefits and harms. Indeed, Singley et al.'s post-deployment concerns about anonymity and reputation [33] echo Miller et al.'s early empirical work on exactly those values.

Empirical studies at IBM did reveal some value issues not considered by Miller et al. In studies of potential users, Huffaker and Lai found that younger workers were more often motivated by reputation, older workers by altruism [20]. Furthermore, the post-deployment survey by Singley et al. revealed that some workers refrained from using the system out of worry that they would bother experts and had nothing to offer in return [33], engaging values of courtesy and reciprocity. Even so, Miller et al. report that the CodeCOOP system is being widely used [26]. Non-users who were interviewed did not cite such concerns as barriers to their use, leading me to speculate that these issues simply did not arise for the CodeCOOP design and context.

Finally, the results of Miller et al.'s survey gave rise to the Value Flows and Dams method, a systematic method for addressing value tradeoffs [26]. Value dams are "technical features or organizational policies that are strongly opposed by even a small set of stakeholders" [26]. Such features ought to be avoided, both for the pragmatic reason of promoting system use and for the ethical reason of respecting the rights and welfare of the minority. Value flows are features and policies that are widely supported though not strictly necessary for the system to accomplish its goals; including such features can make the system even more beneficial to potential users. Considering flows and dams together reveals potential value tensions that arise between or within different stakeholder groups. Pragmatically, the Value Dams and Flows method could help the persuasive technology designer avoid features that will make the system unacceptable to the target audience and include features that make it more attractive—and therefore more persuasive.

Taken together, these two examples demonstrate the worth of VSD's iterative and integrative methodology. By conducting early empirical investigations that are informed by a thorough conceptual analysis of stakeholder groups and the values at stake, the designer can avoid potential minefields in the technical design of the system, and leverage widely held values to make the system more attractive. However, further evaluation of the system should continue to attend to values, both to evaluate user's response to the system with respect to values that were identified early on, and to reveal value concerns that were not anticipated. Thus, explicit and systematic attention to values of moral import throughout the design process can lead to persuasive systems that are both more ethical and more effective.

5. PARTICIPATORY DESIGN

Participatory Design (PD) is a family of theories and methods that involve potential users as full participants in design processes leading to the development of computer systems and computer-based activities [29]. PD grew out of the computerization of Scandinavian workplaces under strong labor rights laws in the late 1970's, stemming from cultural values of human dignity, personal development, quality, and inclusiveness [11]. Many, though not necessarily all, PD researchers and practitioners are motivated in part by a belief in democratizing technology design [29]; woven through the literature is a keen attention to power relationships and a movement to give the vulnerable a voice in technology design. Other motivations for using PD methods include mutual understanding between users and designers, user engagement in the design process, user ownership of the eventual product, and drawing on multiple perspectives to develop innovative solutions. PD has a rich literature developing collaborative design methods such as workshops, storytelling, role-playing, design games, making low-tech models, and cooperative prototyping.

Thus far, little work has considered participatory design of persuasive technology, consisting of only two cases that I discuss in this section. Even so, participatory design methods show promise in addressing challenges of designing persuasive technology. A commonly cited advantage of participatory design is that it promotes a sense of ownership among the technology users. The design process further benefits

from participants' creativity and their knowledge, both explicit and tacit, about the context for technology use. Beyond these usual reasons to employ participatory design, involving potential users in design helps to avoid potential ethical issues with persuasive technology. Potentially vulnerable stakeholders can be engaged in the design process to ensure they have a say in the form the technology takes. Without participatory design, the designer stands outside of the community and intends to change the behavior of community members. With participatory design, the persuasive intent comes in part from community members who want to change the community's behavior from within.

5.1 Case Study: Neighborhood Networks

Coming from the Participatory Design tradition, DiSalvo et al. argue that design itself can be a kind of rhetoric: "the activity of discovering, inventing, and delivering arguments about how we could or should live in the world" [9, p. 41]. Artifacts become rhetorical by their creators' persuasive intentions and how they provoke a response—not only individual changes in behavior or attitude, but dialogue with others. Typically, technology design as a medium for argument is available only to researchers, professional designers, and the technical elite. DiSalvo and colleagues aim to introduce this medium to the public, to engage community members as creators and critics of technology.

DiSalvo et al.'s Neighborhood Networks consists of several community workshops intended to engage residents of Pittsburgh in design for and dialogue about their neighborhoods [9]. The goal is for participants to design "robots"—systems of sensors and actuators—that respond to neighborhood environmental conditions such as noise or air quality, providing persuasive commentary on those conditions.

In 2008, DiSalvo et al. report on the first workshop series, held the previous summer in the Lawrenceville neighborhood [9]. The series comprised four phases in seven, two-hour sessions. The first phase was intended to familiarize participants with the technology and its use in their neighborhoods. The first session took the form of a scavenger hunt, in which small groups of participants were given commercial sensors and asked to respond to prompts such as, "Go someplace you have never gone before and take a sensor reading" [9, p. 43]. Games such as these, central to the participatory design tradition, promote exploration, play, creativity, and critical thinking; they also help participants overcome hesitancy at using the technology [9, 4, 3]. A second session in the familiarization phase introduced the Canary, a simple sensor and actuator platform designed by DiSalvo et al. for the project. Again, the facilitators encouraged an attitude of exploration and play by giving the participants craft materials with which to build simple sculptures that moved in response to sensor inputs.

The next phase focused on concept development. Based on their experiences in the second phase, participants brought objects from home that might be used as part of a robot; for example, a motorized airplane toy might be used to visualize air quality. Afterwards, participants made storyboards, incorporating sketches and words, to show what actions might take place in the environment, what the robots would sense, how the robots would react to what they sense, and how people should respond in turn. Participants worked in an "open studio" format to implement their designs. Researchers no longer guided the activities, but instead supported participants with feedback and technical assistance.

The final workshop was a public event in which participants demonstrated their robots; posters helped support participants' explanations of the robots, how they worked, and their purposes. This final event let participants present their ideas to neighbors, community organizers, and a city planner. The robots provoked dialogue not only about the technology itself, but the participants' reasons for creating what they had: the problems they experienced living in their neighborhood and the kinds of changes they wanted to see. Thus, the robots became objects of rhetoric, a medium for communicating ideas about change in the world.

5.2 Case Study: ADAPT

The Audience Design of Ambient Persuasive Technology (ADAPT) project aimed to engage students at a small liberal arts college as partners in design of a persuasive system to reduce resource consumption in a public location on campus [27]. ADAPT is similar to the Neighborhood Networks project in several respects. Both projects focus on embedding information in the physical world through systems of sensors and actuators; DiSalvo et al. term these "robots" [9] while ADAPT uses the language of ambient displays [35, 24]. ADAPT, too, used a series of summer design workshops to engage participants in designing technology to address a local problem. Similar to Neighborhood Networks, ADAPT's first workshop was an exploratory game in which participants were given tools to explore a space, in this case with the goal of finding places where energy and other resources are consumed. An early workshop introduced the technology; in later workshops, participants were engaged in building mockups with junk and craft materials, and invited to critique early prototypes of several possible designs.

However, where Neighborhood Networks gives individuals and small groups a new medium in which to voice their own concerns, the goal of ADAPT was to implement and install a single, working persuasive computing system. The ultimate goal of Neighborhood Networks was for the participants to use that new medium to persuade other stakeholders in a face-to-face setting to take action on neighborhood problems; the goal of ADAPT was to produce a self-contained, fully automated persuasive computing system informed by participatory methods. This difference in research goals led to significant differences in the technology used and in the balance of control and ownership between participants and researchers. The Canary platform used in the Neighborhood Networks project was designed by the researchers for participants without programming experience; participants control the mapping of inputs to outputs using an on-board menu system [9]. With the intent to build a working system and a lesser focus on participant empowerment, the ADAPT project used the Phidgets toolkit (`www.phidgets.com`) for rapid prototyping of sensor-actuator systems. Although the basic concepts were accessible to participants, the researchers took charge of programming the Phidgets. These differences in control and ownership played out not only in the technology, but also in how power shifted over time in each project. At the end of the Neighborhood Networks project, participants presented their design prototypes; although they were assisted by researchers, the concepts and implementation were substantively their own. By contrast, participants in the ADAPT project fed mainly into the early, concept formation stages of design; the actual implemen-

tation and installation was completed by two student researchers in computer science.

Ultimately, the ADAPT project resulted in an interactive sculpture to occupy a staircase in the science building, intended to attract people from a nearby elevator to the stairs [27]. The system uses LEDs embedded in hand-sized wire sculptures to evoke fireflies along the walls and railings of the stairs; they invite being touched and are responsive to such interaction. Users can also "race the elevator" up and down the stairs; a special animation celebrates users' success. The installation also includes information about the energy use of the particular elevator. The system is currently under evaluation not only with respect to its effectiveness in changing behavior, but also student attitudes toward the display and its community designers.

6. CHALLENGES AND FUTURE WORK

As I have argued and as these initial case studies show, Value Sensitive Design and Participatory Design have great potential for the design of persuasive technology. Such frameworks support the designer in engaging stakeholders and uncovering and addressing ethical issues. Beyond the moral reasons for doing so, there is a pragmatic one as well: systems that address stakeholder values and concerns early on are more likely to be acceptable to users from the beginning. Yet, there is much yet to learn about how to apply these frameworks to the domain of persuasive technology.

For Participatory Design, many of my questions concern the development of appropriate activities for engaging stakeholders in persuasive technology design. Just as games can help to articulate tacit knowledge and values, how can such games help to reveal undesired behaviors, which are often habitual and unnoticed? How can a participatory process incorporate baseline data on current behavior, such as energy consumption or physical activity? How could goal-setting become part of a participatory process? Should participants be introduced to psychological research on persuasion, such as that of Cialdini [6]; if so, when and how? Could techniques such as scenario writing or role playing, for example as conceived by Nathan, et al. [30], be used as part of a participatory process to help stakeholders imagine interactions with and moral reactions to persuasive technology? But, there are also questions concerning the effectiveness of Participatory Design in the persuasive context and its implications for the effectiveness of the technology. For example, what is the relationship of other stakeholders to the persuasive technology that is eventually deployed?

For Value Sensitive Design, I have already sketched how certain questionable methods, namely operant conditioning and surveillance, might relate to values of human autonomy, privacy, accountability, and informed consent. I have proposed the use of the Value Scenarios methods for foreseeing undesired outcomes, and discussed how systems that promote knowledge sharing relate to privacy, awareness, reputation, and trust. Yet, much remains to be learned from applying the VSD framework to a range of persuasive technologies: careful analysis of not only the values implicated by specific persuasive strategies, but also the values implicated by the designer's persuasive intent, such as health, education, and environmental sustainability. Such analysis should lead to criteria that can be used in both empirical and technical evaluations of persuasive technology.

The best way to address these open issues is by engaging in design. For my part, I plan to continue my work on the ADAPT project by conducting Participatory Design activities around environmental conservation at two sites: my institution's EcoHouse and another nearby institution or business. EcoHouse is a college residence occupied by a small group of students who are highly committed to reducing their ecological footprint. EcoHouse has recently installed resource monitoring systems, but these systems require users to sit down at a computer to view data; there is no immediate feedback on resource consumption. Moreover, according to one resident, thus far there has been little discussion among EcoHouse residents regarding how to change their behavior (as opposed to infrastructure) to conserve resources. I plan to work with EcoHouse residents to develop ambient displays that provide feedback on energy or other resource consumption in a form that will help them to adapt their behavior in real-time, incorporating data from energy monitoring systems and possibly other sensor systems in the house. While this work is clearly related to Petersen's recent work on dorm energy use reduction contests [32], a contest is unlikely to be appropriate to this context of use; I will use Participatory Design methods to help develop suitable motivational elements. Work at a second site will lead to more generalizable findings, and also allow me to revise and iterate on the design activities. As a business or other institution, this site is likely to be one in which environmental concerns are important but less salient than in EcoHouse. A key question is how to interweave a Value Sensitive Design perspective with Participatory Design methods.

Beyond my own work, I urge other designers and researchers to consider value sensitive and participatory approaches in resolving their own ethical quandaries in the design of persuasive technology.

Acknowledgments

Thanks to my terrific students, Tim Miller and Pat Rich; to Batya Friedman for suggesting I write this paper; and to Alan Borning and Kate Deibel for their helpful comments on an earlier draft.

7. REFERENCES

[1] B. M. Atkinson. Captology: A critical review. In *Proc. First International Conf. on Persuasive Technology (Persuasive 2006)*, pages 171–182, 2006.

[2] D. Berdichevsky and E. Neuenschwander. Toward an ethics of persuasive technology. *Communications of the ACM*, 43(5):51–58, May 1999.

[3] E. Brandt. Designing exploratory design games: A framework for participation in participatory design? In *Proc. Ninth Participatory Design Conf. (PDC 2006)*, pages 57–66, 2006.

[4] E. Brandt and J. Messeter. Facilitating collaboration through design games. In *Proc. Eighth Participatory Design Conf. (PDC 2004)*, pages 121–131, 2004.

[5] M. Brodie, J. Lai, J. Lenchner, W. Luken, K. Ranganathan, J.-M. Tang, and M. Vukovic. Support services: Persuading employees and customers to do what is in the community's best interest. In *Proc. Second International Conf. on Persuasive Technology (Persuasive 2007)*, pages 121–124, 2007.

[6] R. B. Cialdini. The science of persuasion. *Scientific American*, pages 76–81, February 2001.

[7] S. Consolvo, K. Everitt, I. Smith, and J. A. Landay. Design requirements for technologies that encourage physical activity. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI '06)*, pages 457–466, 2006.

[8] J. Davis. Engaging and informing citizens with household indicators. In *Proc. Hawai'i International Conference on System Sciences (HICSS-41)*, 2008.

[9] C. DiSalvo, I. Nourbakhsh, D. Holstius, A. Akin, and M. Louw. The neighborhood networks project: A case study of critical engagement and creative expression through participatory design. In *Proc. of the Tenth Anniversary Conf. on Participatory Design (PDC 2008)*, pages 41–50, 2008.

[10] K. R. Fleischmann and W. A. Wallace. A covenant with transparency: Opening the black box of models. *Communications of the ACM*, 48(5):93–97, 2005.

[11] C. Floyd, W.-M. Mehl, F.-M. Reisin, G. Schmidt, and G. Wolf. Out of Scandanavia: Alternative approaches to software design and system development. *Human-Computer Interaction*, 4(4):253–350, 1989.

[12] B. Fogg. Persuasive computers: Perspectives and research directions. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI '98)*, pages 225–232, 1998.

[13] B. Fogg. *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann, San Francisco, 2003.

[14] B. Friedman, editor. *Human Values and the Design of Computer Technology*. Cambridge University Press, New York, 1997.

[15] B. Friedman, A. Borning, J. Davis, B. Gill, P. Kahn, T. Kriplean, and P. Lin. Laying the foundations for public participation and value advocacy: Interaction design for a large scale urban simulation. In *Proc. 9th Ann. Int. Conf. on Digital Government Research (dg.o2008)*, Montréal, May 2008.

[16] B. Friedman, D. Howe, and E. Felten. Informed consent in the Mozilla browser: Implementing value sensitive design. In *Proc. Hawaii Int. Conf. on System Science (HICSS-35)*, 2002.

[17] B. Friedman, P. H. Kahn Jr., and A. Borning. Value Sensitive Design and information systems: Three case studies. In *Human-Computer Interaction and Management Information Systems: Foundations*. M.E. Sharpe, Armonk, NY, 2006.

[18] B. Friedman and H. Nissenbaum. Bias in computer systems. *ACM Transactions on Information Systems*, 14(3):330–347, 1996.

[19] C. Graham, P. Benda, S. Howard, J. Balford, N. Bishop, and R. Borland. *"heh—keeps me off the smokes..."*: Probing technology support for personal change. In *OZCHI 2006 Proc.*, pages 221–228, 2006.

[20] D. Huffaker and J. Lai. Movitating online expertise-sharing for informal learning: The influence of age and tenure in knowledge organizations. In *Seventh Annual IEEE International Conf. on Advanced Learning Technologies (ICALT 2007)*, pages 595–599, 2007.

[21] N. Jafarinaimi, J. Forlizzi, A. Hurst, and J. Zimmerman. Breakaway: An ambient display designed to change human behavior. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, pages 1945–1948, 2005.

[22] D. G. Johnson and J. M. Mulvey. Accountability and computer decision systems. *Communications of the ACM*, 38(12):58–64, 1995.

[23] R. R. Johnson. Book reviews: *Persuasive Technology*. *Journal of Business and Technical Communication*, 18(2):251–254, April 2004.

[24] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. Heuristic evaluation of ambient displays. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI 2003)*, pages 169–176, 2003.

[25] R. O. Mason. Morality and models. In W. A. Wallace, editor, *Ethics in Modeling*, pages 183–194. Elsevier Science, Tarrytown, New York, 1994.

[26] J. Miller, B. Friedman, G. Jancke, and B. Gill. Value tensions in design: The value sensitive design, development, and appropriation of a corporation's groupware system. In *Proc. GROUP 2007*, pages 281–290, 2007.

[27] T. M. Miller, P. Rich, and J. Davis. ADAPT: Audience design of ambient persuasive technology. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, April 2009.

[28] L. Millett, B. Friedman, and E. Felten. Cookies and web browser design: Toward realizing informed consent online. In *Proc. SIGCHI conference on Human factors in computing systems (CHI 2001)*, pages 46–52, 2001.

[29] M. J. Muller. Participatory design: The third space in HCI. In J. A. Jacko and A. Sears, editors, *The Human-Computer Interaction Handbook*, chapter 54, pages 1051–1068. Lawrence Erlbaum Associates, Mahwah, NJ, 2003.

[30] L. P. Nathan, B. Friedman, P. Klasnja, S. K. Kane, and J. K. Miller. Envisioning systemic effects on persons and society throughout interactive system design. In *Proc. 7th ACM Conf. on Designing Interactive Systems (DIS '08)*, pages 1–10, 2008.

[31] T. I. Ören, M. S. Elzas, I. Smit, and L. G. Birta. A code of professional ethics for simulationists. In *Proc. Summer Computer Simulation Conf.*, pages 428–433. Society for Computer Simulation, 2002.

[32] J. E. Petersen, V. Shunturov, K. Janda, G. Platt, and K. Weinberger. Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives. *Int. Journal of Sustainability in Higher Ed.*, 8:16–33, 2007.

[33] K. Singley, J. Lai, L. Kuang, and J.-M. Tang. BlueReach: Harnessing synchronous chat to support expertise sharing in a large organization. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, pages 2001–2008, 2008.

[34] M. F. Walji and J. Zhang. Human-centered design of persuasive appointment reminders. In *Proc. Hawaii Int. Conf. on System Sciences (HICSS 41)*, 2008.

[35] C. Wisneski, H. Ishii, A. Dahley, M. Gorbet, S. Brave, B. Ullmer, and P. Yarin. Ambient displays: Turning architectural space into an interface between people and digital information. In *Proc. First International Workshop on Cooperative Buildings (CoBuild '98)*, 1998.