H. Gellersen et al. (eds.), ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work, 18-22 September 2005, Paris, France, 449–468.

Informing Public Deliberation: Value Sensitive Design of Indicators for a Large-Scale Urban Simulation

Alan Borning*, Batya Friedman†, Janet Davis*, and Peyina Lin†
*Department of Computer Science & Engineering; †The Information School
University of Washington
Seattle, Washington, USA
{borning, batya, jlnd, pl3}@u.washington.edu

Abstract. We investigate informing public deliberation regarding major land use and transportation decisions with the results from a sophisticated computer simulation of urban development. Our specific focus is on indicators that portray key results from the simulations. Our design addresses a number of challenges, including responding to the values and interests of diverse stakeholders, making documentation ready-to-hand, and balancing the value of fairness with presenting a diverse set of advocacy positions. We use Value Sensitive Design as our theory and design methodology; our theoretical framework also draws on Habermas's theories of legitimation and communicative action. Our work contributes to CSCW as an example of designing a system for effective use in an environment with multiple stakeholders who have fundamental disagreements, and we conclude by drawing lessons for other environments with these characteristics.

Introduction

Public deliberation and debate over major issues, at local, regional, and national levels, plays a central role in democratic society. We investigate informing such deliberation with results from computer simulations, to help stakeholders understand the long-term consequences of different choices. Facilitating more informed decisions is one side of the coin, but another is supporting the legitimacy of the *process* by which the decisions are made. Our research is guided by the Value Sensitive Design theory and methodology (Friedman, 1997; Friedman, Howe, and Felten, 2002; Friedman, Kahn, and Borning, in press), an approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process. This exploration is grounded in investigations of UrbanSim (Waddell et al., 2003; Waddell and Ulfarsson, 2005), a system for simulating the development of urban areas. UrbanSim is designed to inform public deliberation and debate around major decisions regarding land use and transportation by projecting the long-term impacts of different alternatives. However, the techniques and lessons learned should be applicable to other uses of simulation to support public deliberation, and more generally to enhancing the legitimacy of applications of complex computer systems in the public sphere.

One important characteristic of this problem domain is that it centers on public deliberation and decision-making involving multiple stakeholders. Another is that there are often long-standing disagreements, both on particular projects or legislation, and on the overall approach to urban land use and transportation. These are often rooted in fundamental value conflicts among the stakeholders about issues such as the environment, economic growth, or equity. A third is that the decision-making process can be informed by using modeling and simulation to help reveal the long-term consequences of alternative choices, and that such models can also raise issues themselves, such as concerns regarding “black box” simulations or the input assumptions. From the beginning, our designs have been shaped by these characteristics—in particular the context of a dispute-filled environment, in contrast to more cooperative work environments. These characteristics are shared by other important domains, such as long-range budget forecasts, and environmental issues such as global warming, or biodiversity.

CSCW is concerned with understanding the needs of people who do cooperative work and designing systems to support them. Our work contributes to CSCW as an example of designing a system for effective use in an environment in which participants have fundamental disagreements. That conflict is inherent in cooperative work is well recognized, for even though the core of cooperative work is interdependence in work, this is “by no means necessarily harmonious” (Schmidt and Bannon, 1992, p. 8), and “successful cooperation depends on how conflict is handled” (Easterbrook, 1993, p. 3). Yet, the urban planning context requires a different view of conflict. Conflict between stakeholder groups is often what Pace (1990) calls “competitive conflict” among “entrenched” participants: “[t]his may occur where participants have opposing basic beliefs, values or principles which they believe must be mutually exclusive” (Easterbrook, 1993, p. 25). Competitive conflict may also be seen in groups’ formative stages or within functional teams around power relationships.

Much of the work on group decision-making and deliberation seeks to structure interactions between participants. For example, Decision Support Systems focus on the provision of a decision model (Kraemer and King, 1986). Collaboration tool-based Group Support Systems are intended to support and structure group deliberation (Davison and Briggs, 2000), as are “argumentative” or “discursive” information systems such as Issue-Based Information Systems (Isenmann and Reuter, 1997). In contrast, here we seek to inform rather than to structure deliberation.

In this paper, we first provide an overview of UrbanSim and the role of indicators in presenting its results. We then describe our theoretical framework, including our methodological framework of Value Sensitive Design, and our conceptualization of legitimation and transparency, drawing on Habermas’s theories of communicative action. The bulk of the paper presents the iterative design and formative evaluation of a component of UrbanSim: an Indicator Browser that integrates live Technical Documentation with a chooser for selecting indicators to use to summarize key results from UrbanSim simulations. We conclude with a discussion of directions for future work and implications for other domains that feature multiple stakeholders with strongly held, conflicting views.

UrbanSim Overview

In many regions in the United States, there is great concern about traffic congestion, resource consumption, pollution, loss of open space, lack of sustainability, and 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. Nor are these problems confined to the U.S. In urban regions in Europe, for example, there are often significant disagreements regarding the balance between spending on auto-oriented facilities, public transit, and bicycles; transportation taxes (e.g., on petrol, or fees for taking autos into the city center); or how best to move toward sustainable development.

For example, a regional planning agency might be considering alternate approaches to adding transport capacity, such as a major upgrade to the rail system vs. a new ring motorway around the urban core. To help compare these alternatives and their long-term impacts, the agency might use UrbanSim to simulate the development of the region for the next 30 years under alternative plans, in particular the interactions between transportation and land use, along with their environmental impacts. In assessing the long-term effects of building the ring motorway, we need to consider not only what portion of the current transportation demand can be accommodated by the proposed motorway, but also its effects on land development in the long term around the urban fringe—the motorway might induce additional auto-oriented development in its vicinity, which would in turn add additional traffic, filling the motorway and perhaps creating yet more demand for road capacity. Predicting the future is of course a risky business. In this example, unknown factors that could significantly affect the long-term outcome include the price of oil, possible breakthroughs in technology, or unexpected major shifts in population. Yet we do have to make decisions now, with the information we have. As E.F. Schumaker (1973, p. 240) observed, "The future cannot be forecast, but it can be explored."

To date, UrbanSim has been applied experimentally in the U.S. in metropolitan regions around Eugene/Springfield, Oregon; Seattle, Washington; Honolulu, Hawaii; Salt Lake City, Utah; Houston, Texas; and Phoenix, Arizona; with other applications in process. Internationally, it is being applied experimentally in Paris, France; Tel Aviv, Israel, and elsewhere. It played a significant role in helping settle a lawsuit in Utah regarding a major freeway construction project out-out-court (Waddell and Borning, 2004); the first major use in a public planning process is scheduled to take place in the Puget Sound (Seattle) region beginning in summer 2005. The system is Open Source software, under the GNU Public License, and freely available for download from www.urbansim.org. The system continues to evolve, with the addition of improved and new models.

In urban planning, indicators (Gallopín, 1997; Hart, 1999) are often used to monitor changes in a region with respect to specific attributes of concern. In UrbanSim, simulation results can be presented using the same set of selected indicators for all the policy alternatives being considered, thus aiding the assessment and comparison of different scenarios. For example, suppose that we are interested in fostering compact, walkable, more densely populated neighborhoods within the urban area, and curbing low-density, auto-oriented development ("sprawl"). In the urban planning literature, population density is regarded as one of the key indicators of the character of development (e.g., dense urban, low-density suburban, rural, etc.). We can then monitor population density to understand current trends, and also use UrbanSim to assess and compare the impacts of different policies on population density 30 years in the future. In addition, modelers use UrbanSim indicators diagnostically, to learn about the system's internal operation, to help assess whether it is operating correctly, and to debug problems. In the work reported here, we are concerned with both evaluative and diagnostic uses.

Theoretical Framework: Value Sensitive Design

Our research in this area is guided by the Value Sensitive Design theory and methodology, a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process. Key features of the methodology are its interactional perspective, tripartite methodology, and emphasis on direct and indirect stakeholders.

Value Sensitive Design is an interactional theory: values are viewed neither as inscribed into technology nor as simply transmitted by social forces. Rather, people and social systems affect technological development, and technologies shape (but do not rigidly determine) individual behavior and social systems.

Value Sensitive Design employs a tripartite methodology, consisting of conceptual, empirical, and technical investigations. These investigations are applied iteratively and integratively, with results from new investigations building on and integrating earlier ones. *Conceptual investigations* comprise philosophically informed analyses of the central constructs and issues under investigation. For example, how does the philosophical literature conceptualize certain values and provide criteria for their assessment and implementation? How should we engage in trade-offs among competing values in the design, implementation, and use of information systems? *Empirical investigations* focus on the human response to the technical artifact, and on the larger social context in which the technology is situated. The entire range of quantitative and qualitative methods used in social science research may be applicable. *Technical investigations* focus on the design and performance of the technology itself. Technical investigations can involve either retrospective analyses of existing technologies or the design of new technical mechanisms and systems. The work reported here represents a snapshot of this iterative process: we present our conceptual, empirical, and technical investigations so far, but (as described in the Future Work section) intend to build on these in additional work.

A third key aspect of Value Sensitive Design is its focus on both direct and indirect stakeholders. The direct stakeholders here are the urban modelers and technical planners who use UrbanSim and manipulate its results. The indirect stakeholders are those who don't use the system directly, but who are affected by it. In the case of UrbanSim, the indirect stakeholders include all the residents of the region being modeled, as well as residents of nearby regions.

Early in our conceptual investigations in the present project, we made a sharp distinction between *explicitly supported values* (i.e., ones that we explicitly want to support in the simulation) and *stakeholder values* (i.e., ones that are important to some but not necessarily all of the stakeholders). Next, we committed to several key moral values to support explicitly: fairness and more specifically freedom from bias (Friedman and Nissenbaum, 1996), representativeness, accountability, and support for a democratic society. In turn, as part of supporting a democratic society, we decided that the system 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. Note that explicitly supported values are *not* the same as designer values—they are subjected to a principled analysis of arguments for their inclusion rather than simply being a matter of personal preference. We identified comprehensibility, and subsequently legitimation and transparency, as key instrumental values.

Legitimation: A Habermasian Perspective

UrbanSim's legitimacy is crucial for its effective use as part of the urban planning process. Stakeholders who do not see the use of UrbanSim as legitimate may never accept decisions that are informed by its use, and may disengage from discourse about urban planning, reducing the diversity of stakeholders present at the table and undermining democratic participation. If stakeholders who do not see UrbanSim as legitimate do choose to stay at the table, their constant questioning of simulation results may detract from discourse about what really matters in the outcome of adopting a course of action.

Our conceptualization of legitimation—its central role in the political process, and what allows a political process to be legitimate—draws primarily on the work of Jürgen Habermas (1979, 1984). The legitimation of an urban planning process depends on a huge number of factors. The modeling software forms only one small part, and even the best-designed system could be used in a process lacking in legitimacy. Since most of these factors are out of our control, in this work we concern ourselves with the *legitimation potential* of the modeling system, rather than the legitimation of the entire process in which it plays a part.

Communicative action plays a key role in legitimation potential. Habermas defines communicative action as speech in which all participants aim towards mutual understanding, without manipulative or strategic actions. In communicative action, each utterance implicitly raises four validity claims: to the comprehensibility of the utterance, to the truth of its propositional content, to the truthfulness of the expression of the speaker's intent, and to the rightness and appropriateness of the utterance with respect to existing norms and values. UrbanSim is just one voice in public discourse about urban planning. It does not dictate the truth; rather, it informs a process of coming to an understanding. As it is used in the course of deliberation, information from and about UrbanSim will raise the four validity claims of communicative action. To provide legitimation potential for the use of UrbanSim, we as designers should do our best to ensure these claims are well grounded. First, the information UrbanSim provides should be comprehensible to the range of stakeholders. Second, UrbanSim's models and results should be a reasonable representation of reality. Third, UrbanSim should be transparent with respect to its inner workings and design, so that stakeholders can see that the model and its results are truthfully represented in the deliberation. Fourth, UrbanSim is cast in the role of a source of relatively neutral, technical information in a highly political process. To rightly fulfill this role, and in the interest of fairness to all stakeholders, UrbanSim should provide information that is as unbiased as possible. The information provided should be appropriate and relevant to the policy context.

Those who have access to information such as that provided by UrbanSim have a power advantage in discourse. In the interest of permitting an equal agreement, as many stakeholders as possible should have access to UrbanSim. Many different presentations may be required so that results can be comprehended by stakeholders with differing expertise and accepted by stakeholders with differing norms and values. While the Technical Documentation is intended primarily for modelers and planners, our Indicator Perspectives mechanism, which lets different organizations present perspectives on how UrbanSim output should be used in making policy decisions, is intended for a wide range of interested stakeholders.

Though Habermas has been criticized (sometimes strongly), for our purposes there is much of value here. Indeed, we embrace critiques such as that of Nancy Fraser (1992), who argues that the ideal of the public sphere must be reconstructed to permit the participation of all. According to Fraser, even after everyone is formally licensed to participate in the public sphere, informal barriers such as that of differing communication styles remain. These barriers can be reduced through a multiplicity of publics that give members of subordinated groups safer venues in which to find their voice, so that they can better articulate and defend their interests in the larger public sphere. The Indicator Perspectives mechanism supports multiple publics in that it allows members of particular groups to formulate positions in discourse amongst themselves and then articulate those positions to the larger public.

Transparency and Comprehensibility

The term "transparency" appears in contexts of human-computer interaction, modeling, and public policy, all of which have relevance to UrbanSim. In both human-computer interaction (e.g., Herlocker et al., 2000) and modeling (e.g., Lee, 1973; Fleischmann and Wallace, 2005), "transparency" is used to designate the opposite of a "black box" system, which hides all information beyond its inputs and outputs. In the public policy literature, the term "transparency" is widely used to designate mechanisms for public disclosure of information. Finel and Lord (2000) capture the notion of transparency as a glass box in emphasizing the visibility of the internal characteristics of a government. For simulation models, transparency provides evidence of the system's "truthfulness": that the output reflects the true behavior of the models and is not strategically manipulated. However, a simple "glass box" notion of transparency is insufficient. It is important to make the purpose and assumptions of the system apparent so that stakeholders can assess when its assumptions do not hold or its purpose is incompatible with the goal of the deliberation. Furthermore, Value Sensitive Design leads us to consider transparency for both direct and indirect stakeholders, who will have differing expertise with respect to urban planning, simulation, computer systems, and the region in which UrbanSim is applied. Therefore, transparency is needed at a number of levels—in the reports read by elected officials and the public, in documentation about simulation outputs, in model specifications, and in the availability and comprehensibility of the simulation code itself.

The Design Problem: Challenges with Indicators

We turn now to the design problem that is the focus of the current paper: how can we create an interaction design around indicators for UrbanSim that will provide improved functionality, support stakeholder values, enhance the transparency of the system, and contribute to the system's legitimation? When we began our work, the code to produce indicator output from UrbanSim was intertwined with the simulation code itself, and adding a new indicator was not straightforward. No single list of the implemented indicators existed, and no single place contained the definitions of the indicators or other details that would be needed by modelers and planners working with UrbanSim. There was no easy mechanism for ensuring that indicator documentation was current, including documentation for how indicators were computed. And none of the above information was ready-to-hand (Friedman, Howe, and Felten, 2002; Winograd and Flores, 1986), that is, easy to access in the course of interacting with UrbanSim. With the design and development of the Indicator Browser, we set out to remedy this situation in a way that would help to support the process of legitimation through increased access to and transparency of the indicators. Specifically, we set out to address the following design challenges:

(1) *Fragmentation of indicator information*, in many different sources.
(2) *Lack of ready-to-hand indicator information.*
(3) *Diverse sources and competing definitions for indicators.*
(4) *Difficulty of comprehending indicator information.*
(5) *Difficulty of inspecting and understanding how indicators are computed.*
(6) *Sometimes outdated or inaccurate information.*
(7) *Difficulty of adding and modifying indicators (and corresponding documentation), due in part to the system architecture.*
(8) *Concerns regarding perception of bias in the indicator information*, including what information is provided about the indicators and how they are organized and presented to the user.
(9) *Potentially inadequate representation of stakeholder values*, including a cogent argument for why a given indicator is important and relevant for assessing a particular policy.

We hypothesized that the transparency of the system would be directly enhanced by addressing the first six design challenges. Moreover, we believed that stakeholder representation could be better supported with mechanisms to easily add new indicators. While at the start it was unclear how much progress we could make on any of the first seven design challenges, from our perspective there was little controversy that making progress on any of these would be beneficial.

The last two design goals—that of addressing perceptions of bias and of supporting specific stakeholder value representation—provided a greater challenge, in that they represented a tension between the competing goals of neutrality and value advocacy. In this case, what we sought to make transparent was the purpose of information: when information was of a more neutral flavor and when it clearly represented a specific stakeholder perspective.

The Design Process: The Indicator Browser

In this section we describe our iterative Value Sensitive Design process around the development and informal formative evaluation of the Indicator Browser. Our purpose is to convey how we thought through the value implications of our work and how those analyses impacted our design work. We highlight the integrative nature of our design work, moving among conceptual analyses of transparency, legitimation, representation, and freedom from bias, technical development of the actual Indicator Browser design and implementation, and empirical investigations in the form of informal (and eventually more formal) formative evaluations.

Prototype 1: Envisioning the Indicator Browser

The first problems we addressed concerned information fragmentation, the lack of ready-to-hand information, and balancing tensions between neutrality and value commitments. Our first prototype (Prototype 1) was sketched on a whiteboard (Figure 1) and shortly thereafter developed in MICROSOFT ACCESS (Figure 2). This version divided the screen into two parts: the top part showing the specific indicators a user had selected and the bottom part showing the available indicators to select from. In addition, it grouped the indicators into eight overarching categories and showed the number of indicators selected from each category. The idea was to make visibly salient to users which categories were well represented by any given indicator selection, and which categories less so. We also envisioned a system that would allow users to click on the name of an indicator to bring up ready-to-hand information about that indicator, as well as sample output, though these features were not implemented until Prototypes 2 and 3.

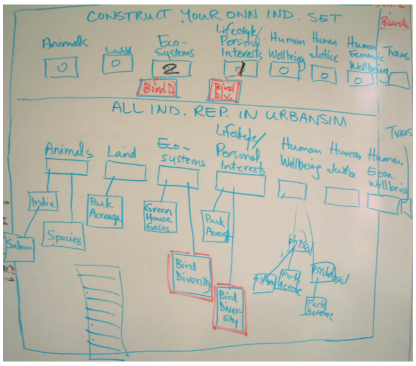

Figure 1. Whiteboard sketch of Prototype 1, showing "Construct Your Own Indicator Set" (top) and "All Indicators Represented in UrbanSim" (bottom), grouped by categories.

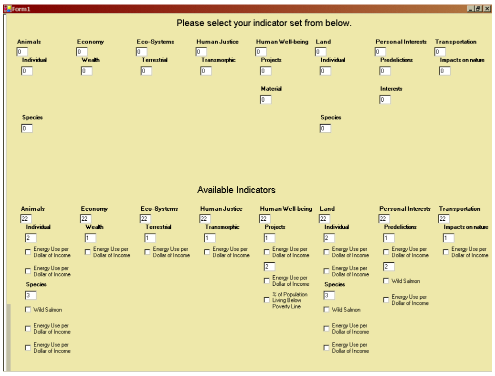

Figure 2. Screen shot of Prototype 1, showing the users' selected indicator set with the number of selected indicators per category (top) and the "Available Indicators" (bottom), grouped by categories.

Prototype 2: Refining the Indicator Browser and Developing the Technical Documentation

The initial sketch of the Indicator Browser realized in Prototype 1 led naturally to the need for two key developments: (a) a change of platform to a web-based implementation that could be readily connected to the working UrbanSim simulation and (b) ready-to-hand Technical Documentation for each of the individual indicators. The former development would allow for a close coupling of the Indicator Browser with the running models and position us to develop live documentation for the indicators; the latter development would increase transparency and comprehensibility of the indicators by providing easy access at the time of use to accurate, useful information about each indicator. With Prototype 2 we set out to design and implement these changes.

At this stage key disagreements arose within the design team regarding categorization schemes that might be perceived as biased and as a result undermine the system's legitimation. Our discussions here were extensive, lasting many months, and nearly bogged down the development of the Indicator Browser. To help move past this log-jam, in Prototype 2 we implemented more than one categorization scheme and put the selection of the categorization scheme into the user's hands. Figure 3 shows one scheme; clicking on the tabs at the top provide a display of the indicators in alternative schemes.

As shown in Figures 3 and 4, Prototype 2 is implemented as a series of web pages. Figure 3 shows the main Indicator Browser page (similar to Prototype 1) with the list of all available indicators now on the left (categorized by Economics, Environment, and Social) and the indicators selected by the user on the right. In

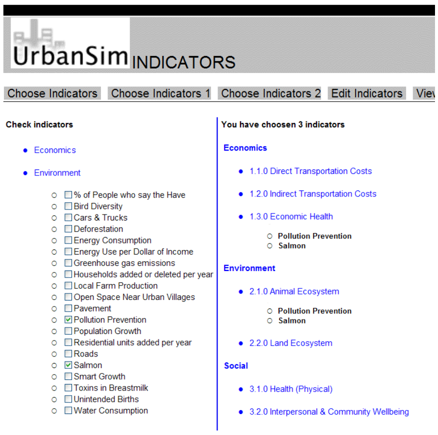

Figure 3. Screen shot of Prototype 2, showing all indicators (left) and user selected indicators (right), categorized as Economics, Environment, or Social. The "Choose Indicators" tabs at the top allow the user to choose among different categorization schemes for the indicators.

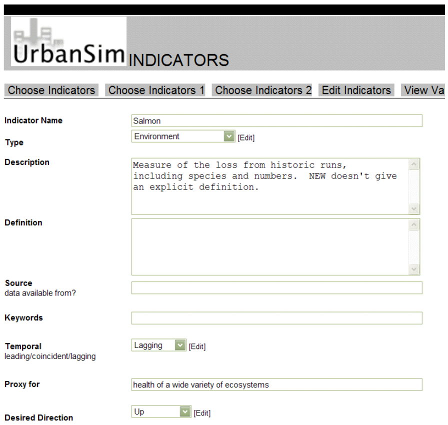

Figure 4. Screen shot of Prototype 2, showing the Technical Documentation about the "Salmon" indicator. Sections shown include indicator name, type, description, definition, source, keywords, temporal relationship, proxy for other indicators, and desired direction of change.

addition, Technical Documentation was created for each indicator. Figure 4 shows sample documentation for the "Salmon" indicator for the Pacific Northwest. Once reasonable Technical Documentation had been developed for the 40 implemented indicators, we began to engage in formative evaluation to refine the content, organization, and presentation of this information.

Informal Formative Evaluation and Iterative Design of Prototype 2

To test our design intuitions, we conducted a series of informal-formative-evaluation/redesign cycles of the Technical Documentation with nine participants, five with a modeling background and four with a policy background. Within each evaluation/redesign cycle, the participant was asked to think aloud while browsing the Technical Documentation in the presence of a facilitator, who made note of the participant's comments and suggestions. We also asked the participant questions about particular documentation elements. Following each evaluation session, changes were made in quick iteration so that each subsequent participant engaged with a slightly improved version of Prototype 2.

Taken as a whole, the formative evaluation guided our redesign to better achieve our design goals. Much of the strongest feedback we received was with respect to neutrality. Early versions of Prototype 2 included a section for the desired direction of the indicator, which we thought would be useful in the context of decision making. However, the information in this section reflected widespread disagreement about the desired direction for many indicators, and some participants indicated that even the section's name conveyed bias. We also experimented with designating indicators as primarily diagnostic or primarily evaluative. Several participants pointed out to us that some indicators we had designated as diagnostic (e.g., "Acres of Developable Land") would in fact be of policy interest to some stakeholders (in this case, real estate developers). Based on this feedback, we eliminated this distinction; instead we included a prominent comment in a new "Interpreting Results" section for the few indicators that report on simulation artifacts and thus are not at all appropriate for evaluating policies.

Prototype 3: The Indicator Browser with Live Technical Documentation and Indicator Perspectives

With the current version of the Indicator Browser (Prototype 3), we completed the work of connecting the web-based Indicator Browser to the live UrbanSim simulation, refined the Technical Documentation in response to the informal formative evaluation reported above, and developed a means for stakeholder groups to advocate for particular value perspectives set apart from the system's Technical Documentation. The decision to create a new separate area within the Indicator Browser for organizations to express their views emerged in response to the need to balance competing requirements for neutrality (the Technical Documentation) and value advocacy (the Indicator Perspectives). In this section, we describe each of these components and highlight how they contribute to our goals for transparency and comprehensibility.

Live Technical Documentation

With an eye toward providing useful, "ready-to-hand," comprehensible information about each indicator as well as minimizing perceptions of bias, we refined the Technical Documentation to include the following sections, as shown in Figure 5: (1) indicator name; (2) definition of the indicator; (3) advice for interpreting indicator results; (4) units of measurement and precision of the results; (5) related indicators; (6) a specification of how the indicator can be computed; (7) any known limitations of the indicator; (8) how the indicator relates to the simulation models; (8) the indicator's source and evolution, as well as examples of its use; (10) the SQL code that is used to compute the indicator from databases of simulation results; and (11) input and expected output for a test to verify that the SQL code computes the indicator results correctly.

This Technical Documentation is "live" in that the SQL code and tests are extracted directly from the code-base each time they are displayed. This guarantees that what the user reads in the Technical Documentation is current.

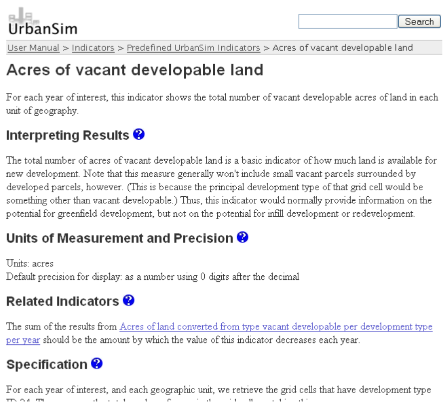

Figure 5. Screen shot of Prototype 3, showing the current Technical Documentation for a particular indicator, in this case "Acres of vacant developable land."

Figure 6. Screen shot of Prototype 3, showing the Indicator Perspective for Northwest Environment Watch.

Moreover, the Technical Documentation is easily updated and extended. And, in keeping with our design goal to create an underlying architecture that can incorporate new indicators readily (Freeman-Benson and Borning, 2003), Technical Documentation can be easily added to the system as new indicators are implemented. Thus we are able to support the extensibility of indicators in UrbanSim, not only technically, but from the user perspective as well.

Indicator Perspectives

As much as possible, the Technical Documentation is intentionally neutral, yet the planning process is rife with strong opinions and perspectives. How then might stakeholders use the indicators to represent and express their strongly held opinions? Here we have taken an approach—Indicator Perspectives—that allows stakeholders to tell a story and advocate particular values and criteria for evaluating outcomes. The Indicator Perspectives position organizations to present their own perspectives on which indicators are most important for evaluating policy alternatives, and how those indicators should be interpreted. We believe that these perspectives will be useful to stakeholders and decision makers because the organizations have well thought-out positions and can present them clearly and coherently. In contrast to the Technical Documentation, which is intended to cover all the indicators in a fairly neutral way, Indicator Perspectives can focus on a small set of indicators in a potentially opinionated way. Indicator Perspectives are intended to provoke thought and public deliberation, as well as to give groups a venue in which to state their positions.

We are currently in the early stages of developing Indicator Perspectives. We have partnered with three local organizations to construct perspectives for the initial prototype: a government agency (King County Budget Office, which publishes the King County Benchmark Reports), a business association (Washington Association of Realtors), and an environmental group (Northwest Environment Watch). In keeping with our explicitly supported value of representativeness, we choose initial partners who cover a wide range of views. Later, we plan to provide opportunities for involvement to all who are interested, actively soliciting partners as needed to help ensure coverage of the political and policy space. Figure 6 shows one prototype perspective, based on the Cascadia Scorecard, a monitoring program developed by Northwest Environment Watch.

Evaluating the Indicator Browser with Urban Planners

At this formative stage of our design process, we sought to systematically evaluate several key aspects of the Indicator Browser and its component pieces. While we hypothesized that we had solved key aspects of the information fragmentation problem (both in terms of consolidating information and making it ready-to-hand) and would positively impact task performance (e.g., comprehension and evaluation of indicators), we had not tested our redesign work. We also hypothesized that design features such as providing live SQL code, limitations of the indicators, and test case information would increase comprehensibility and transparency of the indicators. There were also unresolved design issues concerning the categorization of indicators, tools for on-demand testing of indicators, incorporating region-specific documentation, and maintaining the visibility of unused indicators. To answer these and related questions, we conducted a small user study focusing on the Technical Documentation with urban planners interested in UrbanSim, who constitute the primary audience for the Technical Documentation. (Since we were in the early stages of developing the Indicator Perspectives, we decided to evaluate this component with a broader range of stakeholders at a later date.)

Participants and Method

Eight current or prospective UrbanSim users (2 women; 6 men) participated. Participants were recruited at an UrbanSim user group meeting and had at least some urban planning experience (Range: 1 to 22 years; $M = 10.5$ years).

Each participant was engaged in a semi-structured interview for approximately one hour and fifteen minutes. The value-oriented interview questions and tasks drew in structure on prior research (Friedman, Kahn, and Hagman, 2004). The first group of questions explored the participants' current work practices, including their estimates of the time it would take to complete various tasks related to indicators and the number of sources they would need to consult. The participants were then asked to identify values and policies important to land use and transportation in their own regions, and to record these on cards. Following a demonstration of the Indicator Browser, participants were asked to perform short tasks using the Technical Documentation (e.g., defining an indicator in their own words, describing the relationship between two indicators, identifying three indicators to assess a particular concern). Participants were then asked about design tradeoffs with respect to ten current or future design decisions for the Technical Documentation. Each design trade-off was presented in terms of two alternate views with the rationale tied to transparency and comprehensibility supporting each view, e.g., for live SQL code:

> View 1: One person told me that including the SQL code in the documentation is helpful. Reading the code helps you to know what's really going on when the indicator is computed. Including the code in the documentation makes it easy to find. It's also easy to compare the code to the definition of the indicator and the specification of how it should be computed. Even if I don't read the code, it's reassuring to see it there and know that it's the actual code that is run to compute the indicator values. It's just more transparent that way.
>
> View 2: Another person told me that including the SQL code in the documentation is not very helpful. Other sections of the documentation, like the definition and the specification, provide all the information you will usually need about how the indicator is defined and how it's computed. The code is lengthy and hard to read compared to these other sections. If you need more information, you can always go find the source code somewhere else.)

Participants were asked to identify the view more like their own. Finally, participants were asked to identify indicators that would be informative for evaluating scenarios with respect to the values or policies they identified at the beginning of the interview. A subsequent telephone interview was conducted with seven of the eight participants to supplement incomplete work practice data.

All interviews were audio recorded for later transcription. A coding manual was developed to code evaluations and responses to content questions. Data were coded by two independent coders trained in the coding manual. Intercoder reliability was assessed through testing Cohen's kappa at the □ = .05 significance level; all tests were statistically significant, with k = .74 - .94 depending on question type. For the short tasks, time to complete each task was recorded, as well as whether the participant consulted the Technical Documentation. A domain expert assessed whether each task was completed correctly.

Results

For the task performance questions, participants required much less time to complete each of the four tasks using the Indicator Browser than with their traditional work practices (Table I). For 26 of the 27 tasks (96%) for which we have both estimates and data on task performance, the time it took the subject to complete the task using the Indicator Browser was less than the estimated time that they gave based on current work practice. The overall median estimated time was 20-60 minutes, while the actual median time to complete the tasks was only

Task	Median Time to Complete Task: Estimated Time in Current Practice	Median Time to Complete Task: Actual Time using Indicator Browser	Actual performance less than estimated current practice?	Wilcoxon *p*-value
1. Define Nonresidential Square Feet	10-20 min.	1.7 min.	6 out of 7 (86%)	.014*
2. Discuss relationship between Residential and Household Density	10-20 min.	2.8 min.	7 out of 7 (100%)	.009*
3. Find three indicators of economic growth	20-60 min.	2.4 min.	7 out of 7 (100%)	.009*
4. Explain what Jobs-Housing Balance says about commute times	20-60 min.	2.0 min.	6 out of 6 (100%)	.022*

Table I. Task Performance. An asterisk (*) denotes significant differences at the p = .05 level between estimated time to complete tasks in current practice and actual time using the Indicator Browser.

Feature	Description	p-value
Categorize Indicators	In the Indicator Browser opening screen, the indicators are grouped according to some categorization scheme (the specific scheme is not specified) rather than alphabetized.	.035*
Interpreting Results	In the Technical Documentation, the Interpreting Results section provides advice for understanding what the indicators signify and how to use them to answer different kinds of questions.	.008*
Limitations	In the Technical Documentation, the Limitations section provides information about pitfalls in using the indicator as well as when to avoid using the indicator altogether.	.008*
Live SQL Code	In the Technical Documentation, the live SQL code section provides access to the code used to compute the indicator; the "live" SQL code is extracted from the code-base at display time.	.008*
Test Cases	In the Technical Documentation, the Test Cases provide SQL code that is run to test the indicator.	.008*
Do Not Display All Categories	In the Indicator Browser opening screen, the categories are always displayed even if no indicators from a given category are selected.	.060
Do Not Distinguish Diag. & Eval. Indicators	In the Indicator Browser opening screen, indicators are designated as evaluative or diagnostic.	.164
Do Not Include Test-On-Demand	In the Technical Documentation, the Test-on-Demand section allows the user to run the indicator Test Cases from the web.	.125
Layered Documentation	In the Technical Documentation, include region-specific information along side of the UrbanSim software documentation.	.453
Specific Categorization Scheme	In the Indicator Browser opening screen, choose between two competing categorization schemes, one based on non-expert conceptions and one on the urban planning literature.	1.000
Comprehensive List of Indicators	In the Indicator Browser opening screen, provide a comprehensive list of indicators for stakeholder values about regional land use, transportation, and environmental impacts, including those that UrbanSim may not yet support.	1.000

Table II. Design Tradeoffs. An asterisk (*) denotes features that were preferred by a significant number of participants at the p = .05 level.

2.1 minutes, indicating a substantial improvement. Participants were also asked to estimate the number of sources they would need to consult in their current work practice to complete the tasks. For each of the four tasks, the median was two to three sources. By comparison, each participant who successfully completed the task in the study did so using only one source, the Indicator Browser. Of the 31 tasks performed by the 8 subjects, 25 were completed successfully (81%).

The design tradeoff questions, in which participants were asked to select one of two views, were analyzed using a binomial test. Table II provides a description of the specific design features and summarizes the quantitative results. As shown in Table II, nearly all participants preferred categorizing the indicators in the Indicator Browser opening screen ($p = .035$) as well as including sections for each indicator in the Technical Documentation on Interpreting Results ($p = .008$), Limitations ($p = .008$), Live SQL Code ($p = .004$), and Test Cases ($p = .008$).

Regarding linking values and policies with indicators, participants generated a total of 31 values and policies related to urban planning in their regions (Range: 3-4; M = 4), 18 of which were within UrbanSim's current scope of land use, real estate, employment, and demographic indicators (Range: 1-4; M = 2). Participants were later asked to identify indicators that they believed would inform discussion of those values and policies. Of the values that were within UrbanSim's current scope, 6 were not considered due to time constraints, and 2 were deemed by the participant to be unsuited to the use of indicators. For 9 of the 10 values and policies (90%) for which participants attempted to identify indicators, participants were able to find informative indicators.

Discussion

Taken together, the results on task performance and design trade-offs indicate that much is working here to support comprehensibility and transparency of indicators in UrbanSim. In particular, the results on current work practice and task performance with the Indicator Browser provide strong support that the current design—with cohesive ready-to-hand Technical Documentation—has made progress toward addressing the problem of information fragmentation. The significant positive assessments from the design trade-off questions confirm our decisions to include the Interpreting Results, Limitations, Live SQL, and Test Case sections to improve the policy relevance and transparency of the Technical Documentation. In addition, results provide some support that we were successful in providing indicators that are appropriate to values and polices that are important to the stakeholders in the participants' regions.

Other results point to directions for future design. For example, though all participants supported some form of categorization, there was no consensus on which scheme to use. These results suggest a need for further investigation of categorization schemes, as well as design of a ready-to-hand mechanism for choosing among multiple categorizations along the lines of our earlier implementation in Prototype 2. Also, further work is needed on how to handle regional information.

Lessons Learned and Future Directions

We believe the research reported here represents a successful application of Value Sensitive Design theory and methodology to the problem of informing public deliberation using sophisticated computer models. In this section we reflect on the lessons learned thus far and their broader implications.

First, the distinction between explicitly supported values and stakeholder values has held up well throughout our research. Because explicitly supported values are subject to a principled analysis of arguments for their inclusion, this distinction provides a strong response to the concern that the system simply reflects the personal values of the designers. We recommend making this same distinction in the conceptual analysis in other CSCW domains that feature multiple stakeholders with strongly held, divergent values.

Second, the identification of legitimation potential as an instrumental value has allowed us to draw on the rich theoretical work of Jürgen Habermas as well as that of some of his critics, and provided a useful way to reconceptualize the organization of some of our original explicitly supported instrumental values. Habermas's theory of communicative action in turn leads to a set of testable design goals (comprehensibility, accuracy, transparency, relevance, and freedom from bias). For UrbanSim, legitimation potential is in support of the moral value of fostering a democratic society, but an analogous move could be made in other CSCW domains in which the legitimacy of the use of a system may be in question.

Third, for complex systems such as UrbanSim, minimizing information fragmentation and providing ready-to-hand documentation can go some distance toward the goals of comprehensibility, transparency, and relevance. Specific techniques that we used, and that could be gainfully employed in other contexts, include live code and tests (integrated with the documentation), as well as integrated discussion of limitations and how to interpret results.

Finally, to address the tension between possible perceptions of bias on the one hand, and value advocacy and engaging citizens in the democratic process on the other, we provide both relatively neutral technical information and also a diverse spectrum of advocacy positions, distinct but interlinked. As discussed earlier, work on the Indicator Perspectives is in the early stages. However, we are optimistic that this work will unfold to provide additional lessons for balancing value advocacy with freedom from bias in other contested domains.

Planned work includes improving the comprehensibility of the information for the range of stakeholders, integrating the Indicator Perspectives more closely with the other components of the Indicator Browser, and providing better support for comparison and discussion of different perspectives. Finally, we intend to deploy and evaluate the Indicator Browser and Indicator Perspectives in a real decision-making context; our plan is to do so in the upcoming revision of VISION 2020, the regional growth and transportation plan in the Puget Sound (Seattle) area.

In conclusion, we have provided a snapshot of an ongoing research project on informing public deliberation with the results from sophisticated simulations. We believe that the lessons learned so far—in particular, regarding the use of Value Sensitive Design, the strong distinction between explicitly supported values and stakeholder values, the focus on legitimation, minimizing information fragmentation and providing ready-to-hand-documentation, and techniques for balancing value advocacy and neutrality—can be valuable in other CSCW domains involving multiple stakeholders with strongly held, divergent values.

Acknowledgments

Thanks to Peter Kahn and Paul Waddell for many useful discussions of the underlying ideas; to Ann Hendrickson, Bjorn Freeman-Benson, Jon Fuchs, Casey Huggins, Charles Naumer, and David Socha for help with the design and implementation of different versions of this system; to Stephanie Collett, Shireen Deboo, Nicole Gustine, and Jessica Miller for help with the formative evaluation; to Lance Bennett for discussion about Habermas; to Brian Gill for his assistance with statistics; and to our user study participants. This research has been supported in part by the U.S. National Science Foundation under Grant Nos. EIA-0121326 and IIS-0325035, and in part by gifts from IBM and Google.

References

Davison, R. M., and Briggs, R. O. (2000): 'GSS for Presentation Support', *CACM*, vol. 43, no. 9, pp. 91-97.

Easterbrook, S. M., Beck, E. E., Goodlet, J. S., Plowman, L., Sharples, M., & Wood, C. C. (1993). 'A Survey of Empirical Studies of Conflict', in S. M. Easterbrook (Ed.), *CSCW: Cooperation or Conflict,* Springer-Verlag, London, pp. 1-68.

Finel, B. I. and Lord, K. M. (2000): 'The Surprising Logic of Transparency', in B. I. Finel and K. M. Lord (eds.): *Power and Conflict in the Age of Transparency*, Palgrave, New York.

Fleischmann, K. R. and Wallace, W. A. (2005): 'A Covenant with Transparency: Opening the Black Box of Models', *CACM*, col. 48, no. 5, pp. 93-97.

Fraser, N. (1992): 'Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy', in C. Calhoun (ed.): *Habermas and the Public Sphere,* MIT Press.

Freeman-Benson, B., and Borning, A. (2003): 'YP and Urban Simulation: Applying an Agile Programming Methodology in a Politically Tempestuous Domain', *Proceedings of the 2003 Agile Development Conference*, Salt Lake City, Utah, June 2003.

Friedman, B. (ed.) (1997): *Human Values and the Design of Computer Technology*, Cambridge University Press and CSLI: New York, NY and Stanford, CA.

Friedman B., Howe, D. C., Felten, E. W. (2002): 'Informed Consent in the Mozilla Browser: Implementing Value Sensitive Design', in *Proceedings of the 35th Hawaii International Conference on System Science.*

Friedman, B., Kahn, P. and Borning, A. (in press): 'Value Sensitive Design and Information Systems', to appear in P. Zhang & D. Galletta (eds.): *Human-Computer Interaction in Management Information Systems: Foundations*, M.E. Sharpe, New York.

Friedman, B., Kahn, P. H., Jr., and Hagman, Jr. (2004). 'The Watcher and The Watched: Social Judgments about Privacy in a Public Place', in *Online Proceedings of CHI Fringe 2004.* Vienna, Austria.

Friedman, B. and Nissenbaum, H. (1996): 'Bias in Computer Systems', A*CM Transactions on Information Systems,* vol. 14, no. 3, July 1996, pp. 330-347.

Gallopín, G. C. (1997). 'Indicators and Their Use: Information for Decision-making', in B. Moldan & S. Billharz (Eds.), *Sustainability Indicators: Report of the Project on Indicators of Sustainable Development* (pp. 13 - 27): John Wiley & Sons Ltd.

Habermas, J. (1979): *Communication and the Evolution of Society*, T. McCarthy (trans.), Beacon Press, Boston.

Habermas, J. (1984): *The Theory of Communicative Action*, Vol. 1, T. McCarthy (trans.), Beacon Press, Boston.

Hart, M. (1999). *Guide to Sustainable Community Indicators* (2nd Edition). North Andover, MA.

Herlocker, J. L., Konstan, J. A., and Riedl, J. (2000): 'Explaining Collaborative Filtering Recommendations', in *Proceedings of CSCW 2000.*

Isenmann, S. and Reuter, W. D. (1997): 'IBIS—a Convincing Concept…But a Lousy Instrument?', in *Proceedings of ACM Designing Interactive Systems.*

Kraemer, K. L. and King, J. L. (1986): 'Computer-Based Systems for Cooperative Work and Group Decisionmaking: Status of Use and Problems in Development', in *Proceedings of CSCW 1986.*

Lee, D. B. Jr. (1973): 'Requiem for Large-Scale Models', *Journal of the American Institute of Planners*, vol. 39, no. 3, May 1973, pp. 163-178.

Pace, R. C. (1990): 'Personalized and Depersonalized Conflict in Small Group Discussions: An Examination of Differentiation', *Small Group,* vol. 21, no. 1, pp. 79-96.

Schmidt, K. and Bannon, L. (1992): 'Taking CSCW Seriously: Supporting Articulation Work', *Computer Supported Cooperative Work: An International Journal,* vol. 1, no. 1 pp. 7-40.

Schumacher, E.F. (1973): *Small is Beautiful: Economics as if People Mattered*, Harper, New York.

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*, vol. 3, no. 1, pp. 43-67.

Waddell, P., and Borning, A. (2004): 'A Case Study in Digital Government: Developing and Applying UrbanSim, a System for Simulating Urban Land Use, Transportation, and Environmental Impacts', *Social Science Computer Review*, vol. 22 no. 1, pp. 37-51.

Waddell, P., and Ulfarsson, G. (2005): 'Introduction to Urban Simulation: Design and Development of Operational Models', forthcoming in P. Stopher, K. Button, K. Haynes and D. Hensher (eds.): *Handbook in Transport, Volume 5: Transport Geography and Spatial Systems*, Pergamon Press.

Winograd, T. and Flores, F. (1986): *Understanding Computers and Cognition: A New Foundation for Design,* Norwood, New Jersey, Ablex.