

Ambient Information Systems: Evaluation in Two Paradigms

Zachary Pousman, John Stasko
School of Interactive Computing and GVU
Georgia Institute of Technology
Atlanta, GA 30332
{zach | stasko} @cc.gatech.edu

ABSTRACT

We take two paradigms for information systems development, functionalism and social relativism, and apply their assumptions to the evaluation of ambient information systems. Ambient information Systems research, we posit, comes from two distinct paradigms and this has confounded a single evaluation framework from emerging. Instead, different groups of researchers have specific (if implicit) philosophical commitments about people and the social world, and these commitments have led to two research paradigms. We explain our view of these evaluation paradigms, and note a single area of focus for each evaluation framework. For functionalist evaluations, the questions circle around what to measure (and how to measure). For social relativist evaluations, the questions are practical but also theoretical; are we even asking the right questions? If so, how could we answer them?

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

GENERAL TERMS

Ambient Information Systems, Human Factors., Longitudinal Evaluation, User Study, Design

INTRODUCTION

Every technology product has, if one looks deeply, a delicate and complicated relationship with its users, and the environments in which users live and work. Ambient Information Systems, those systems that display important but not critical information on aesthetically pleasing and subtle displays, are no different [12]. In fact, when compared to systems that users employ for work tasks, they almost certainly have this character. Evaluating Ambient Information Systems (AISs) is a process of explaining how often an AIS is used, how much it improves the life of users, and how the system is appropriated into the rhythms and practices of users. AISs are

built by designers from different research paradigms. Designers subscribe, either implicitly or explicitly to a set of practices and commitments about the world that make up a research paradigm [7]. These research paradigms shape the ways in which designers of systems view the world, including their assumptions about the nature of the social world and the goals of technology intervention. The paradigms also influence the evaluation questions that are asked, the metrics by which success will be judged, as well as the methods they will employ in measuring these phenomena.

In some research areas, all or nearly all the practitioners have a consistent world view, and have wide agreement on their assumptions. This leads them to agreement on what matters for design, as well as what counts as success (success metrics), and evaluation methods. In contrast, AIS design and evaluation is complicated because the devices and systems that are produced span two differing research paradigms. We believe that some research teams have particular commitments and beliefs on the status of the world, while others subscribe to a different set of beliefs. Let us explain a bit what we mean by the idea of a research paradigm, and specifically the two paradigms that we feel that researchers in our community use in their work. Paradigms, as we the authors understand them, can be implicit or explicit, so some researchers may not have consciously or publicly joined a paradigm, though we claim the paradigm still impinges on their work, the systems they build, and the ways that they evaluate those systems. We follow Hirschheim and Klein, who explained four different theoretical frameworks for the entire space of information systems development [7].

Their analysis of the philosophical and social science commitments of research communities mirrors Burrell and Morgan, sociologists who sought to explain the commitments of Sociological research [1]. The first spectrum represents the reality of the social world and whether (and how) it can be understood. On one extreme are the realists (empiricists), who believe in a stable, perceptible, and explainable social world, and who are scientific in their explanations. At the other extreme are those theorists who believe the social world to be socially constructed, and use anti-rationalist explanations (at its most extreme, where the social world could be described as imaginary, this is philosophical Solipsism). A second spectrum concerns explanations of behavior in social experiences. We feel that this spectrum of control and regulation versus transformative change (even revolutionary

change) is not as applicable to the design community for AISs, so we will not dwell on it here. Four quadrants fall out of the crossing of the two spectra. They are: Functionalism, Social Relativism, Radical Humanism, and Radical Structuralism.

We posit that two paradigms, Functionalism and Social relativism are relevant to researchers who build and study Ambient Information Displays. Functionalism is the earliest paradigm for system's research, and has guided the work of engineers and developers. Hirschheim and Klein claim that Social Relativism came about as a reaction to the functionalist research paradigm. Social relativism contains phenomenological and hermenutic philosophical positions, where the social world is understandable only by investigating the everyday "felt lif" of a particular group, and the ways that they use signs and sign-systems to negotiate these meanings. Hirschheim and Klein, as well as Burrell and Morgan before them, claim that the four paradigms are mutually exclusive and irreducible; they claim that the two paradigms cannot be joined in a synthesis that would combine the best elements from each. This may make it impossible to create a single paradigm for design and evaluation of AISs.

AISs appear to us to be built by research groups from the Functionalist paradigm or the Social Relativist paradigm. This paper will work to explain how these two paradigms address the messy issues surrounding evaluation of AISs. Many researchers have already noted the difficulty in designing, running, and analyzing the results from evaluations (as we note in the next section). Our hope here is to use the opposing research paradigms to hone in on a single evaluation framework under each paradigm. The evaluation frameworks that we propose for each paradigm does not change the basic shape of evaluations of ambient systems. Successful evaluations are by and large longitudinal in their character, taking a view of technology that unfolds over weeks and months, not minutes. These evaluations predominantly take place in situ, situated in authentic work and home environments. As opposed to short laboratory-based studies, these long-term and situated evaluations of technology aim for rich narratives and ecologically valid results, no matter the particular paradigm chosen.

RESEARCH PARADIGMS

First we should explain in greater detail the two paradigms that we find operating in the space of ambient information systems. These paradigms are "the most fundamental set of assumptions adopted by a professional community" [7]. The assumptions in a paradigm concern the very nature of what a particular community (and therefore a researcher in that community) can see, because the assumptions filter the social world based on concepts and objects that are governed by the paradigm. Said another way, joining a paradigm requires a commitment from members about ontology, the kinds of entities that exist in the social world. Paradigms also concern epistemology, the way that knowledge can be gathered. Further paradigms make claims about methodology, how research is done research, and the nature of human beings, whether they are mostly volitional (free to chose

their actions) or constrained by external forces and events. Paradigms, as we have already hinted, are potentially unconscious and part of the unquestioned background of research practice.

The paradigm of Functionalism is the oldest and still most common research paradigm in Computer Science and Information Systems development [7]. Functionalism states that the social world is real, stable, teachable, and rationally understandable. The paradigm commits to the existence of mental phenomena in the lives of people. People are rational beings with free will, and their fundamental social driver the coordination, regulation, and control of social situations (this is opposed to understanding the social conflict as primarily a class struggle of owners of capital versus laborers, an assumption by those who subscribe to a more structuralist paradigm). System design in this paradigm can be thought of as "instrumental reasoning" and the role that the developer plays is "systems expert." The expert moves through the analysis process by decomposing goals into hierarchical tasks, and then proceeds to develop a system that supports "rational organizational operation and effective and efficient project management." [7]. The goal of the developer is to increase efficiency and effectiveness. Hirschheim and Klein claim this is the oldest and most ingrained paradigm in IS development, and we feel that it has continued to be important to HCI, ubicomp, and ambient information system research. The paradigm of Social Relativism is a more interpretivist research paradigm. Researchers who subscribe to this paradigm have different commitments, different understandings about the world, and their roles as analysts and developers reflect their beliefs.

The Social Relativist paradigm takes reality as socially constructed, and therefore potentially different for different people. The social world is complex, full of traditions, social conventions, and cultural norms, which are not purely rational. People are embedded in the social world, which is ever-changing, and in some real sense continually reinvented by individuals and groups. On this shifting ground, social practices may never be fully rationalizable, nor is rationality the only way to understand human behavior. The systems developer who takes on this paradigm is more accurately described as a "facilitator" and attempts to tease out some (potentially limited) order by a process of sense-making [7]. The goal of the developer is to help a given population achieve consensus on their shared understanding of a confusing (and potentially unknowable) reality.

We will now move the discussion to the view that each paradigm brings to evaluation, in the context of AISs. We are not judging these paradigms on their relative value, and our aim is not to show one paradigm as a better fit for AIS design and evaluation. However, a clear understanding of the paradigms coupled with how those paradigms find expression in the evaluations of ambient information systems may help the evaluator ask clearer, more germane questions, design better studies, and, we hope, get more compelling results from those studies.

FUNCTIONALIST AIS EVALUATION

Inside the functionalist research paradigm, AIS evaluation has a particular character. We summarize our conception of that character here. An AIS is designed to convey information in a way that is calm and not disruptive or distracting. An AIS is designed to be helpful for the important but not critical information in our lives, and to increase our awareness. In this paradigm, researchers know what they are looking for, and the only puzzle is to find it. A successful AIS would increase a user's perception of her awareness of the information displayed. This increase in awareness would improve her quality of life and decrease her feelings of stress and anxiety (or at least those stresses caused by not having this information at the ready). The question then becomes: *How do we measure these constructs in user's heads?*

The evaluation of an AIS is most often done in an authentic situation, such as in the home or office setting. Evaluations are often longitudinal, so that the evaluators can be sure that results are not skewed by the "novelty effects" early on. Surveys and other self-reporting measures can be used to gather feedback on measures concerning efficiency and effectiveness of the system. The evaluators measure the sense of usefulness, anxiety, distraction, and perceived time savings. An example of this evaluation technique is the evaluation of the InfoCanvas by Stasko *et. al.* [14] Stasko and his team deployed a consolidated information display, the InfoCanvas, into eight participant's offices for a period of one month. They interviewed participants with a pre-intervention, mid-point, and end-point feedback sessions. These sessions consisted of a battery of Likert-scale questions surveys to determine if and how much user's perceptions were changing over the course of the study. The sessions also contained more open-ended questions, which sought to gather narratives of situations where the system came in handy. These qualitative data points were intended to bolster the claims from the quantitative measures and their trends throughout the study. Other studies that are in this paradigm, and that therefore follow this evaluation framework (but in a laboratory setting) are McCrickard, *et. al.*'s and Matthews, *et. al.*'s studies of design for the periphery of computer screens [11, 9].

The evaluation framework of functionalism requires that members of the research community who subscribe to this set of commitments come to a clear set of metrics. That work is, from our reading in the field, ongoing, but certain metrics have gathering consensus. We list some candidates here, with a short discussion:

- **Heightened Awareness** An AIS will make information available for quick and opportunistic glances
- **Time Saving** The hope is that AIS are convenient for users, saving them time in their routines
- **Anxiety / Stress** An AIS should decrease the anxiety and stress that comes from having to actively find and monitor information sources. If a user feels little stress from monitoring a particular data source, then an AIS that includes this information is of little value.

- **Distraction of presentation** An AIS should have more payoff in heightened awareness and time savings than it has detriment being distracting. We note that distraction can mitigate over time, so a evaluation should last long enough for the system to become routine.
- **Learnability of mappings/representations** An AIS is often evaluated on the success of the mapping between data and presentation. A system that remains hard to decode is not a success.

Is this a complete list, or should other important metrics be added? The functionalist paradigm for AIS research hopes to find a consistent set of metrics through which developers can make claims about the success or failure of their ambient systems. The list of metrics should also be measurable in a consistent (and generalizable) way. New methods for getting at these perceptions can be developed. Also, observational data could potentially be used to get past the perception to the actual participant behavior for metrics like counts of "glances per day" (see Shen for an interesting attempt at this [13]) and "distracted by display."

SOCIAL RELATIVIST AIS EVALUATION

For researchers in a social relativist framework, AIS evaluation is conceived differently. In light of their philosophical commitments and orientation, the AIS intervention needs to be studied in light of complicated and harder to measure topics like work practices and rhythms, appropriation, and adoption of systems into the lives of participants. These evaluations are of similar setup when compared to functionalist evaluations. They are long-term, *in situ* studies by and large, but the focus of the evaluation, driven by the philosophical commitments, is quite different. Researchers in this paradigm might believe that it is impossible to look inside the heads of users and participants. Or, slightly less categorically, they might believe it impossible to know precisely if and when they've found something. The most one might be able to say is that the evaluator might have been able to note moments of reflection and contemplation in the lives of participants. Carter, *et. al.* note "The central problem facing developers of peripheral displays is that metrics for success are not well defined. One [expert interview] participant summarized the issue saying 'most technology that is out there is about maximizing efficiency' that is often not the case with peripheral displays, causing designers to 'reevaluate [standard] systems of evaluation'." [2] If, as Carter claims, AISs are not about maximizing efficiency, then what are they about?

One place that many researchers have looked for answers to this question is phenomenology, a philosophy that operates firmly in the social relativist paradigm. Hallnas and Redstrom characterize their design and evaluation of an AIS for weather data for multiple cities in a way that is not on the previous list of metrics that matter for evaluation [6]. Instead, they draw on phenomenological philosophy, and propose the notion of *presence* as key to analyzing and evaluating AISs. They contrast presence with *use* or *function*. Hallnas and Redstrom note "Having encountered problems such as how to evaluate a certain design and how to describe what

constitutes good design in these areas, we came to question the relevance of some of the basic assumptions in human-computer interaction.” [6] McCarthy and Wright, though they are focusing on system evaluation in a wider domain than just AIS, claim that the felt life and emotional quality of system interaction are of key importance [10].

The methods for evaluation in a social relativist paradigm are narrative and ethnographic in their main approach. These techniques rely on the evaluator being, himself or herself, the tool with which data is gathered. The evaluator, attempting to stay open to various kinds and modes of appropriation attempts to find the ways (if any) that participants are making the AIS part of their lives. These evaluations are hardly ever comparative in their execution, and can be conceived of not experimental interventions. However, the goal is not generalizable results, but, instead, a consensus between the users and the developers themselves around the use of the AIS. Techniques beyond ethnography that are used by researchers in this paradigm are bringing users into the evaluation, making the evaluation part of the experience of using the system [8]. Other approaches that are undertaken are probes (cultural and technical), experience sampling methods, and other methods to spark and stimulate reflection by users on the system under evaluation [3, 4]. AISs can be evaluated in the social relativist paradigm from multiple angles and perspectives in a kind of triangulation. This can even stretch toward mixed methods research, which blends quantitative (historical or demographic), qualitative, and participatory evaluation to find multiple corroborative data points from which a holistic portrait of use emerges.

The social relativist paradigm for AIS evaluation hopes to answer the question of how exactly, to what degree, and to what ends, users integrate AISs into their lives. Answers to this question are at a level at which only descriptive statistics can be gathered, and instead, rich narratives chart the appropriation of the technology into the lives of users. Gaver’s work with the History Tablecloth epitomizes, for us, this approach [5]. The ongoing questions for this research paradigm consist of work to extend the paradigm, to define its course, and to refine the methods utilized by the community.

CONCLUSION AND FINAL QUESTIONS

In this position paper we have proposed that there are two different paradigms of research at work in AIS design and evaluation. We note that they may be mutually exclusive, governed by philosophical commitment of the researcher more than pragmatic or practical strategy. As such, we attempted to give our views on the paradigms, the entities they focus on, and the way that these paradigms find expression in the evaluation of systems. Now we take up two specific areas, one in each paradigm, where open work remains for the researchers who are committed to that paradigm. We propose a more focused and nuanced vision of evaluation questions for practitioners in each paradigm.

For the Functionalists: *Is the list of measures for evaluation complete? If not, what is missing? How can we achieve agreement on how to study these measures as user’s engage*

with ambient information systems?

For the Social Relativists: *Are the concerns we mention for evaluation ever going to achieve closure? Will phenomenology, a philosophical position that many researchers in the area appeal to, be able to provide deep insight? What specific evaluation questions seem germane given this philosophical framework? How can they be answered?*

REFERENCES

1. G. Burrell and G. Morgan. *Sociological Paradigms and Organizational Analysis*. Ashgate, 1979.
2. S. Carter, J. Mankoff, S. Klemmer, and T. Matthews. Exiting the cleanroom: On ecological validity and ubiquitous computing. *Human Computer Interaction (HCIJ)*, IN PRESS, 2006.
3. B. Gaver, T. Dunne, and E. Pacenti. Design: Cultural probes. *interactions*, 6(1):21–29, 1999.
4. W. Gaver, A. Boucher, S. Pennington, and B. Walker. Evaluating technologies for ludic engagement. In *Proceedings of CHI, Extended Abstracts*, 2005.
5. W. Gaver, J. Bowers, A. Boucher, A. Law, S. Pennington, and N. Villar. The history tablecloth: illuminating domestic activity. *Proceedings of DIS, Designing Interactive systems*, pages ??–??, 2006.
6. L. Hallnäs and J. Redström. From use to presence: on the expressions and aesthetics of everyday computational things. *ACM Transactions on Computer-Human Interaction*, 9(2):106–124, 2002.
7. R. Hirschheim and H. K. Klein. Four paradigms of information systems development. *Communications of the ACM*, 32(10):1119–1215, October 1989.
8. K. Isbister, K. Hook, M. Sharp, and J. Laaksolahti. The sensual evaluation instrument: developing an affective evaluation tool. In *CHI ’06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1163–1172, New York, NY, USA, 2006. ACM Press.
9. T. Matthews, M. Czerwinski, G. Robertson, and D. Tan. Clipping lists and change borders: improving multitasking efficiency with peripheral information design. In *CHI ’06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 989–998, New York, NY, USA, 2006. ACM Press.
10. J. McCarthy and P. Wright. *Technology as Experience*. MIT Press, 2004.
11. D. S. McCrickard, R. Catrambone, C. M. Chewar, and J. Stasko. Establishing tradeoffs that leverage attention for utility: Empirically evaluating information display in notification systems. In *International Journal of Human-Computer Studies*, volume 8, pages 547–582, 2003.

12. Z. Pousman and J. Stasko. A taxonomy of ambient information systems: Four patterns of design. In *Proceedings of Advanced Visual Interfaces*, pages 67–74, May 2006.
13. X. Shen, A. V. Moere, and P. Eades. Find: An intrusive evaluation of peripheral display. In *GRAPHITE '05: Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 289–292, New York, NY, USA, 2005. ACM Press.
14. J. Stasko, D. McColgin, T. Miller, C. Plaue, and Z. Pousman. Evaluating the infocanvas peripheral awareness system: A longitudinal, in situ study. Technical Report GIT-GVU-05-08, Georgia Institute of Technology, March 2005.