

Double Binds and Double Blinds: Evaluation Tactics in Critically Oriented HCI

Vera Khovanskaya
Information Science
Cornell University
vdk9@cornell.edu

Eric P. S. Baumer
Communication and
Information Science
Cornell University
ericpsb@cornell.edu

Phoebe Sengers
Information Science and
Science & Technology Studies
Cornell University
pjs54@cornell.edu

ABSTRACT

Critically oriented researchers within Human-Computer Interaction (HCI) have fruitfully intersected design and critical analysis to engage users and designers in reflection on underlying values, assumptions and dominant practices in technology. To successfully integrate this work within the HCI community, critically oriented researchers have tactically engaged with dominant practices within HCI in the design and evaluation of their work. This paper draws attention to the ways that tactical engagement with aspects of HCI evaluation methodology shapes and bears consequences for critically oriented research. We reflect on three of our own experiences evaluating critically oriented designs and trace challenges that we faced to the ways that sensibilities about generalizable knowledge are manifested in HCI evaluation methodology. Drawing from our own experiences, as well as other influential critically oriented design projects in HCI, we articulate some of the trade-offs involved in consciously adopting or not adopting certain normative aspects of HCI evaluation. We argue that some forms of this engagement can hamstring researchers from pursuing their intended research goals and have consequences beyond specific research projects to affect the normative discourse in the field as a whole.

Author Keywords

Critical technical practice; critically oriented HCI, evaluation

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

In 1987, two graduate students at MIT published a paper

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in TimesNewRoman 8 point font. Please do not change or modify the size of this text box.

Every submission will be assigned their own unique DOI string to be included here.

describing a new way to construct intelligent programs in the major annual Artificial Intelligence (AI) conference [2]. This paper became widely influential, spawning a new area of AI research and nine years later named the most important paper in the conference that year. It was not, however, influential in the way that at least one of the authors intended. For Philip Agre, the paper represented not only a breakthrough in the technical development of AI, but also a critical reframing of AI as a research program. As later worked out in *Computation and Human Experience*, for Agre the primary goal of this work had been to critically analyze the limitations of the philosophical conceptualizations undergirding technical work in AI and to offer a new conceptual alternative, embodied in technical language that engineers could understand. This approach Agre termed critical technical practice. Yet, to his frustration, while AI enthusiastically picked up on the technical work, the fundamental critical reconceptualization of human activity embodied in the technology was not picked up; eventually the work was reabsorbed into what Agre saw as business as usual in AI [44].

While critical technical practice as a concept did not have the lasting influence on AI that Agre had hoped, in recent years the idea of deeply coupling technical design work with critical reflection on its conceptualization has been picked up in a neighboring field, Human-Computer Interaction (HCI). For example, within HCI, Dourish cites Agre's notion of critical technical practice as a major inspiration for his landmark work *Where the Action Is* [16, p 4]. Sengers et al.'s articulation of reflective design [41] expands critical technical practice to be used not only to work toward the resolution of specific technical impasses, but as a core method throughout all phases of technology's design and use. The engagement of critical practice within technology design specifically and HCI more broadly has also been spurred through research drawing on other critical traditions. For example, Bardzell, Bardzell, and Blythe integrate criticism and critical theory into design [5,6,7,10]; DiSalvo and Hirsch draw on critical and political traditions in the arts to frame new opportunities and roles for design [15,23]; and Pierce and Löwgren use philosophy of technology and design theory to articulate and execute critical approaches to interaction design [35,36,30]. In short, there is a foment of critical approaches in HCI which aim to push forward HCI research while simultaneously

troubling some of its foundational assumptions. Thus, in contrast to Agre's impact within AI, in HCI critically oriented research has clearly achieved lift-off as a vibrant research program.

In this paper, we take the emergence of a wider body of research integrating critical reflection with technology design as an opportunity to better understand the dynamics and problematics of tactically engaging with HCI in order to expand, reconceive, or otherwise critically reform the field. We argue that the challenges that Agre faced in marrying critical reflection with an ongoing technical discourse that otherwise tends to bracket such questions have not simply disappeared but remain relevant to understanding how researchers can successfully tactically engage HCI. Moreover, HCI adds a significant new complexity to this challenge because of the split audience of HCI work. HCI researchers must engage not only other scholars, but also end users or other participants, who are an obligatory passage point for commonly accepted definitions of what makes a system or approach 'work' in HCI.

Within this set of considerations, we aim to make three main contributions. First, we articulate how the need for tactically engaging with the lingua franca in HCI shapes the nature of our interactions with the participant in ways that can hamstring at least some of researchers' critical intentions. Second, through examining unpublished backstories of three of our own projects, we demonstrate how this happens concretely on the ground, leading to some undesirable practical consequences in relationships with participants despite our largely being able to position the projects as successful from an HCI research point of view. Finally, based on our own work and related published projects, we identify three specific ways in which what counts as 'working' in mainstream HCI discourse manifests in critically oriented HCI discourse in ways that can inadvertently lead to particular forms of breakdown in the relationship between designers and publics.

TACTICALLY ENGAGING HCI

The problematic we explore in this paper is associated with work in HCI that proposes significant alterations to core foundational assumptions of HCI and embodies these alternatives in working systems or methods. Some of this work is explicitly inspired by critical technical practice, some of it comes from other critical traditions, and some is not explicitly framed as 'critical' but still aims to open significant new spaces for HCI by inverting dominant assumptions of what makes for good design or research. In this section, we will describe how tactical engagement with HCI works in several well-known projects and articulate three key attributes of these projects that support effective tactical engagements with HCI. We will argue that these three key attributes lead to a core problematic for critically oriented HCI work: that the need to articulate systems as 'working' within HCI's discursive norms significantly troubles our relationships with participants.

What does tactical engagement with HCI look like?

We will ground our ongoing discussion in this section in three examples of 'working systems' which have been influential in the field and inspirational for our own practice. We chose these particular exemplars as instances where authors begin to "lift the curtain" to reveal some of the tensions in evaluating critical HCI projects.

The first system, *Affector*, is an interactive video installation developed by Phoebe Sengers, Simeon Warner, and Kirsten Boehner [12,40,42]. The system is comprised of a video window between the offices of two friends that enables them to communicate their emotions by systematically filtering the video feed according to sensor readings. *Affector* is presented as a challenge to dominant threads in computer-mediated communication toward greater realism and accuracy in representation of affect by not directly representing emotion in the system and instead emphasizing openness to interpretation. Methodologically, it challenges HCI notions of objectivity and scientific observation by using an 'autobiographical design' method in which the same people are designers, users, and evaluators, working together to iteratively design and develop the system in reaction to their personal experiences.

The second system, *Local Barometer*, is one of three "threshold devices", poetic intermediaries between home and the environment, developed by Gaver's research studio at Goldsmiths [21, 33]. The *Local Barometer* is intended to provide people with a new sense of the "sociocultural texture" around their house. The barometer itself is composed of six small screens which display scrolling images and texts from classified advertisements sourced from the internet. The local wind conditions, as measured by the device, determine which advertisements appear on the screens: the harder the wind blows, the greater the geographic distance from which the ads "travel." Gaver et al. position this work to resist a dominant ethos in technology design towards accuracy and efficiency. Instead, they stress the value of multiple, unresolved narratives in understanding the meanings of technology, including play, ambiguity and interpretation in design [42]. Thus, the evaluation of this system involved collecting and juxtaposing multiple narratives of use, rather than emphasizing a single correct meaning or use of the system.

The third system is an "ultra low cost sensing system" developed by Kuznetsov, Hudson, and Paulos to support air quality activists in detecting particulate pollution [25]. The "lo-tech" sensors can be assembled inexpensively from common paper materials. The authors describe the system and its deployment in a local environmental community, highlighting a series of tradeoffs between expensive and inexpensive sensing systems, usability and generalizability, and the role of inexpensive sensors for encouraging reflection and community action. Through this system, the authors aim to alter the political economy of environmental sensing by making it possible for non-experts to collect,

measure, and reflect on local air quality, rather than requiring them to partner with scientists who would be in charge of the data gathering and interpretation. Kuznetsov et al. test the system in collaboration with a local activist group. This project is part of a longer-term design research program on participatory approaches to environmental data gathering, which has highlighted political issues in the design of environmental sensing systems [4,25,26,27,28].

Based on these examples, we now describe three key attributes of these systems and others like them that support effective, critical engagement with more mainstream HCI. Later, we examine how these strategies interact to shape what it is possible and not possible to say through constructing working systems in HCI.

Engaging the Lingua Franca

One attribute emerging from these examples is that they make their critical interventions by challenging some aspects of the normative discourse while simultaneously engaging, and to a degree upholding, its other elements. Even while each approach highlights the ways it upends one or more significant aspects of standard HCI practice, each also supports the viability of its critical move by making sometimes implicit and sometimes explicit appeals to HCI's established "lingua franca"—the common forms of communication of an intellectual field. In this section, we explore how these three examples do so by looking at what aspects of HCI they queer and what they hold steady.

With Affecter, the authors were not only arguing for a new, non-representational way of looking at affect. They also argued for shifting the evaluation of design practice in HCI from ascertaining whether a design is a success or failure by predetermined metrics to narrating how the system "came to be known or lived as a success or failure" [12, p. 12:24]. They do so by recognizing "multiple, perhaps conflicting interpretations" [42 p. 101] of the design through iterative and thoughtful analysis not only of the system design, but also of the evaluation criteria. As the authors point out, one of the "problematics" [12 p.12:19] their work revealed was that knowledge gained through these approaches "is not always compatible with what is recognized as knowledge production in HCI" [12 p.12:19].

The evaluation of Affecter was explicitly motivated in part by the need to counter perception by the broader HCI community that allowing for multiple possible interpretations rather than a single, correct answer would lead to an "anything goes" mentality. One stated goal was to show that multiply-interpretable systems could still be rigorously evaluated [42]. In an effort to make room for their new approaches and to habilitate them to HCI sensibilities, the authors appropriated and reframed existing HCI methods. In their writing, the authors explicitly aimed to use these methods coupled to a different epistemological stance. For example, they used a diary for the two users to track their emotions and interactions with the system, and

they assigned to their third collaborator the role of an "external evaluator" to provoke some of the reflection and moderate discussions about the system. In other work, Sengers legitimated these practices by arguing that approaches such as Affecter's which address multiple interpretations are still "definable and testable" by HCI standards [42, p.107].

For Gaver's team, there is a similarly strong commitment to presenting an alternative to dominant technology design methods. This work is not framed as 'critical' - Gaver et al. see this work rather as a positively and practically oriented design practice. Nevertheless, it directly challenges some HCI orthodoxies, particularly traditional goals of "usefulness and usability," and aims to open HCI eyes to the possibilities of designing for open-ended playfulness (vs tight functionality) and leveraging ambiguity (vs. establishing clear answers) [18].

This challenge is embodied not only in the goals designed for in these systems, but also in their evaluation. In contradistinction to approaches that aim to establish definitively how a system is taken up in use, the evaluation of the Local Barometer, as well as other domestic technologies such as the Drift Table and the Key Table, focuses on capturing the intervention's multiply interpretable assessments [42]. Gaver and his team study families who are tasked to "live with" the technology by using ethnographic techniques. They also solicit the views of semi-independent commentators, such as journalists and documentary filmmakers, who, like the participants, are not informed of any designerly intentions for the deployment. These methodologies can read as strange to an HCI audience more attuned to laboratory-based 'user tests' conducted with behavioral social-science methodologies, and are probably explicitly intending to be provocatively different. These moves are not naive; as with Affecter, Gaver's explicit intention is to co-opt the documentary method while recouping it for a different epistemological orientation. For his team, the double-blind techniques are a way of capturing the multiply interpretable, to avoid narrowing the scope of the interpretation and to "open up" the design space as much as possible.

In the low-tech sensors case, the challenges to HCI orthodoxy have to do with questions about the framing of the relationship between researchers and participants. Particularly in Aoki et al.'s ethnographic work into the challenges of environmental sensing [4], on which the low-tech sensors work explicitly draws, a key issue is seen in the way activist groups defy usual HCI understandings of fixed, well-defined user populations who are positively oriented to technology and to research. That study identified activist groups' skepticism, directed towards scientists seeking to understand air quality, around who has access to the data collected and who is in charge of its interpretation. This finding lead directly to Kuznetsov et al.'s choice to design air quality sensors that can be easily

interpreted by activists themselves, thus removing scientists from the equation. The design is evaluated through a community workshop in which activists try out the sensors and the researchers interpret the results. The narration of the results blends issues and factors from a standard, apolitical HCI with those that come from the political stance of the researchers, ranging from discussions of ease of use to reflections on the degree to which the sensors allow for end users to feel a sense of ownership and understanding of the data. At the same time, the evaluation is couched in terms drawn from standard HCI methodologies, particularly the form of the focus group. For example, the interpretations provided are the researchers', the participants are anonymized and referred to through generic, interchangeable labels, and the argument is substantiated through empirical evidence in the form of direct quotations from participants.

In all three examples, then, we see critical interventions couched as challenging HCI norms, coupled to aspects that buttress the validity of the challenge with appeal to terms and sensibilities that resonate, or are thought to resonate, with a more mainstream HCI community. Such strategies are sometimes framed as concessions or compromises required to get a work published [29]. That framing, however, does not do justice to the creative challenge constituted by the commitment to engage the mainstream discourse for change. Each of these examples remixes the mainstream and the subversive in a different way in order to establish new grounds. By identifying this dynamic, we wish, in part, to call it out as an explicit and challenging research design problem worthy of sustained attention within the critically oriented HCI community.

At the same time, we need to be aware of the consequences and costs of the specific ways in which the dance between the mainstream and subversion takes place. Agre's experience in AI, described in the introduction, serves as a case in point. In [3 pp. 152-153] Agre describes why it is difficult for critical alternatives to truly redirect a technical field. One of these reasons is that "it is difficult to apply [a] method [embodying a critical alternative] to any real cases without inventing a lot of additional methods as well, since any worthwhile system will require the application of several interlocking methods, and use of the existing methods may distort the novel method back toward the traditional mechanisms and the traditional ways of talking about them." There is thus a risk that the necessary business of leveraging traditional HCI methodologies and sensibilities may limit our ability to shape true critical alternatives. Further, accommodating mainstream sensibilities may make it possible for critical alternatives to be picked up, but at the same time invite re-appropriation into the status quo. Boehner et al., for example, argue that Gaver's methodology of 'cultural probes,' intended as a radical reframing of HCI techniques for assessing use contexts that undermined claims to objectivity, became quickly reframed and "improved upon" by assimilating it to

more recognizably behavioral-scientific forms of interpretation [11].

Still, the risks of being re-assimilated seem attenuated in HCI compared to AI because of the presence of an established community of researchers schooled in critically oriented sensibilities, which likely supports more radical methodological transformations than would be possible if papers were only ever reviewed by researchers committed to mainstream methodologies. But while the risks described by Agre in AI may be less dire in contemporary HCI, the adoption of critical engagement with technology design as part of HCI practice also introduces a significant new element, in the form of users.

Constructing what 'works'

A second attribute emerging from these examples is that they aim to make critical interventions in HCI through embodying them in 'working systems', or, to put it more precisely, systems that can be narrated to the HCI community as 'working.' The critical arguments gain rhetorical force in the community by demonstrating that one can challenge significant norms of HCI practice and still end up with a system that is practical, usable, functional, or otherwise clearly 'good' by standards that are at least somewhat recognizable from mainstream HCI.

This 'work ethic' [3] is also a central feature of Agre's critical technical practice in AI; his systems were intended to give rhetorical force to his philosophical critique by demonstrating that alternatives generated from that critique can be embodied in a working system. In [3], Agre argues that the notions of what it means for a system to 'work' shape the claims it is possible (or not) to make through them. Within HCI, particularly in the last 10-15 years, claims to a 'working system' have been rooted nearly universally, as the previous section might suggest, in processes of 'evaluation', where evaluation almost invariably involves tests with human participants. In order to make the claim that a critical intervention 'works' as a design principle, researchers frequently choose to engage in 'evaluation,' i.e., muster arguments that involve claims based on empirical evidence of how participants understand or interact with a system. Sometimes, as with the low-tech sensors, critical interventions in HCI are embodied in challenging assumptions or norms about use in the design of a system, while those systems are evaluated using fairly standard strategies such as usability tests or focus groups. Other times, as with Affecter and Local Barometer, researchers alter the evaluation methodologies themselves in ways they consider to better align with the critical goals of the project.

In considering these choices in a critically oriented design project, we note that, since the validity of the design argument hinges on a convincing evaluation, there is likely more room to maneuver tactically in the design of a system than in its evaluation. Thus, the question of evaluation is a

lynchpin for understanding how critically engaged projects gain a rhetorical foothold in HCI. At the same time, 'evaluation' must not be framed too narrowly. As we saw with the Affecter, Local Barometer, and Lo-tech Sensing examples unpacked previously, the negotiation of "what works" is present not only during the literal "doing" of the evaluation, but throughout the construction and narration of design and evaluation.

Reframing relationships with users

Given that participants are a central resource for knowledge construction in HCI, a third attribute that emerges from these examples is that in critically oriented HCI not only the technology itself but also relationships between designers, technologies, and their users become material to be critically reframed. All three systems share an interest in challenging traditional relationships between technology interventions and human participants by presenting alternative relationships in the form of working systems. Affecter, for example, is grounded in a critique of how standard affective computing frames the relationship between human emotional experience and its mechanic representation; it also is used to argue that richer notions of emotional experience can be brought into design by violating common notions of objectivity in HCI through folding designers' experiences directly into design. Gaver et al. use the Local Barometer to argue that forms of human experience normally marginalized in design are legitimate design material. Further, the system is based on reframing the relationship between designers and users from designers attempting to control user experience to enabling more open-ended forms of engagement. Low-tech sensors are designed for a politically engaged audience that implicitly challenges traditional HCI conceptions of the discipline as scientific and therefore politically neutral. Explicitly, the sensor project aims to reframe relationships between scientists and activists by putting its users in control of sensing and interpretation and cutting the scientists out of the loop.

More generally, and in contradistinction to the case in AI (at least at the time Agre was working), critical engagement by HCI researchers is oriented not only to critical reflection on the conceptual commitments of the discipline, but also to altering how we imagine and perform relationships between researchers and users. Sometimes this is embodied in a commitment that not only researchers, but also end users, participants, or publics should critically reflect on technology and its relationship to human life [e.g. 41]. This interest in considering and responding to the problematic politics of designers and users deeply resonates with and is partially inspired by the long legacy of participatory design. But, unlike participatory design, in which these reframed relationships are widely considered an inviolable aspect of what it takes to create knowledge, in HCI the commitment to developing alternative relationships to users comes under substantial stress. The next section explains why.

The Problematic

So far in this argument, we have identified three key attributes of projects that embody critical challenges to mainstream HCI in working systems. First, researchers must necessarily engage with the lingua franca to make their critical challenges legible and defensible in the field; they do so by holding some aspects of mainstream methodology steady while violating others. Second, in order to argue for the value of their critical challenge, they frequently engage in some form of 'evaluation' which is based on narrating the results of empirical interactions with human participants. Third, part of the critical project in HCI involves queering the relationships among designers, technology, and users as commonly imagined in the field.

A major resulting problematic, and one we have experienced in our own work, is that *the act of making critically oriented design interventions legible to the HCI community—i.e. tactically engaging with the "lingua franca"—shapes the nature of interactions with participants in ways that can undermine the critical goals of the project.* Keeping in mind that legitimate evaluation is part of the HCI definition of "working," it follows that discursive norms in the HCI community will have an impact on how critically researchers engage with the human participants in their work (i.e. their "users") in order to be able to speak convincingly to the broader HCI community. These simultaneous conversations with the twin audiences of scholars and participants can result in complicated double binds. These double binds arise because critically engaged projects necessarily find themselves grappling with a legacy of evaluation in HCI in order to negotiate what it means for their own systems to "work" in accordance with its discursive norms. And what makes an evaluation 'convincing' in HCI is strongly influenced by its inheritance from the intellectual traditions of psychology, cognitive science, and human factors. This is the complex and somewhat contradictory discursive world in which critically oriented researchers – often coming from humanities, arts, or design traditions to some degree at odds with behavioral social-science epistemologies – must maneuver to make our arguments hold.

In the rest of this paper, we aim to lay out concretely how the way we construct legitimate claims in evaluation shapes the relationships it is possible for us to have with participants. Our goal in this analysis is to look at specific ways that tactical engagement takes place, and the consequences of trade-offs made in that engagement. We will argue that our relationships with users become contorted in particular ways through specific strategies for legitimation inherited from mainstream HCI.

These trade-offs and their consequences are often hard to see, particularly in published papers. This is because those papers are necessarily framing interactions with users in ways that will support claims to legitimacy. The contortions and problematics are outside that frame and perhaps even to

some degree outside researchers' conscious understanding. Next, we will explore the nature of and substantiate this problematic by looking at problems that emerged in three examples from our own work. We will lift the veil from our previously published papers to discuss issues in our relationships with participants that troubled us at the time, but seemed to fall outside the frames of what we could easily discuss in the published work. Later in the paper, we will integrate these experiences with published works to identify discursive tropes which shape engagement with audiences and account for design and evaluation outcomes.

TENSIONS, TRADE-OFFS, AND DOUBLE B(L)INDS

The following three research projects have been published in some capacity in HCI venues and have involved at least one of the authors. Some of these examples explicitly draw from critical technical practice and critical design methodologies, while others are not explicitly “critical” and instead engage in subversion through other forms of unusual HCI practices. What is held in common in these three accounts is the presence of the three attributes described above: engaging with HCI’s lingua franca and with human participants, while simultaneously challenging status quo relationships between researchers, users, and technological systems through “working” alternatives. By examining our own work through these internal accounts we can investigate the ramifications of the interactions between attributes more fully than by reading only the published portions of this critically oriented work.

Designing for Frame Reflection

One of the authors was involved in designing and implementing an interactive visualization tool that leverages computational linguistic analysis to present patterns of language in political blogs and news sources. In contrast to traditional data visualization, the design sought not to provide an overview of *what* is being said but rather *how* things are being said, i.e., to encourage attention to and reflection on how issues are “framed” [17,13]. In addition to lab studies and focus groups [37], this system was deployed in a field study, during which the tool was used for at least 8 weeks by regular readers of political news coverage during the 2012 U.S. election campaign [8]. The evaluation sought to assess the tool's capacity for supporting frame reflection [39], the ways that users integrated tool use with their existing reading practices, and broader issues in how participants interpreted the computational analysis and visualization. So as not to bias their responses, participants were not initially told that the primary intended purpose of the system was to foster frame reflection.

The field study identified a variety of ways that participants used the frame reflection tool, but particularly relevant here is that participants often attempted to use the system as a ‘bias detector.’ Participants would examine specific patterns of language visualized using the tool in an effort to

assess the biases of different sources and sort out opinion from factual news reporting and identify reporterly bias. While instances of frame reflection that were in line with Schön and Rein's [39] description of frame reflection did occur, they were mentioned more rarely.

The first set of tensions arises from participants' insistence on the value of neutral, factual news reporting. Indeed, during controlled experiments and focus group studies that involved explicitly mentioning the concept of framing [37], participants often interpreted “framing” as “spin” or “bias.” The system was seen as a tool for identifying and avoiding such manipulations of objective news reportage. This usage, however, runs contrary to literature on framing, which suggests that there is no such thing as an un-framed fact [17,13]. This distinction between biased and “straight” news, between opinion and fact, led participants to use the system not in a reflective mode but in an evaluative one. Thus, the accounts that participants often gave were in some ways at odds with the frame reflection and the conceptual goals for the paper.

Second, we had hoped the system would help participants identify frames at work, consider the assumptions on which those frames are based, and look at the ramifications of those frames for the perception. However, as prerequisite to engaging in frame reflection, participants would have needed to accept that framing operated in the way described by these theories. That is, in order to engage in frame reflection, our participants would have needed to adopt a meta-perspective that (1) multiple perspectives on (i.e., framings of) reality exist and (2) those different perspectives (framings), generally speaking, are each equally valid. Once this meta-perspective is adopted, a tool such as the one we designed could potentially become quite useful. However, without such a meta-perspective, the visualizations provided by the tool become open to (mis)interpretations that align with a different fundamental take on the world and the nature of (objective) reality.

In retrospect, we experienced a researchers’ double bind: if we told participants about framing, frame reflection, and multiple perspectives, then any evidence of frame reflection we saw in their accounts would be more likely attributable to those statements than to their experiences with the tool per se. However, when we kept such information from them, as was the case here, then participants were left with little scaffolding for the kind of thinking the tool was meant to engender. This double bind did not result in the research becoming entirely incapacitated, as some participants were able to come to an understanding of framing, but the process was hamstrung by our desire to evaluate the validity of the system for supporting frame reflection.

Challenging Self-Optimization

In this project, two of the authors were involved in the development and deployment of a reflective, critically oriented personal informatics tool [24]. This tool was

designed to inspire reflection not on peoples' own behaviors, as personal informatics systems usually do, but instead, on the infrastructures underlying the gathering and presentation of personal data and the narrow modes of engagement that traditional personal informatics systems promote. We encouraged this critical reflection by gathering users' web browsing data and displaying it using different design strategies: for example, we hoped to encourage reverse engineering of data gathering algorithms from our participants by using purposeful malfunction in the visualization. Because we were arguing for the role of critical design in challenging the status quo of self-optimization narratives in personal informatics, and because we wanted to show that our design strategies could potentially be employed toward similar means in other personal-data contexts, our evaluation needed to convincingly illustrate the ways that the designs themselves provoked the specific sorts of reflection we intended.

In order to evaluate our system, we recruited participants for a study about reflecting on web browsing data and about unexpected approaches to visualizing personal information. While this was not directly deceptive, it did not directly call out the critical agenda in our work. During the interviews with our participants, we found ourselves in a complicated tension. Participants picked up on the strangeness of our system's design but were also trying to use the system as one would use a traditional personal informatics tool. Often, this interpretation of the tool would happen in the context of "critical" conversations where participants speculated on the norms, limitations, and assumptions in personal informatics systems. While some participants expressed skepticism toward personal informatics systems and the self-optimizing values they espouse, it is undeniable, too, that some of those same participants were optimistic about tracking and imagined using the tool to "improve" their web browsing habits.

We discovered a complex negotiation between the participants and the interviewers that evolved in response to users' conflicted interpretations of the system. We wanted to get responses from participants during the interview without revealing our own motivations, but simultaneously felt tempted to explain our critical stance in order to have an informed conversation about the limitations of personal informatics. The way this played out in practice was that the former behavior manifested during the "formal" interview, while the latter discussion occurred after the interviewer began "debriefing" the participant about the goals of the study. This ad-hoc negotiation with traditional HCI protocols allowed researchers to pick up on threads previous articulated by participants earlier in the interview while maintaining the "hands-off" ethos of not biasing our participants (at least not initially). The negotiation also demonstrates the ways that our research goals shaped the ways we engaged with the participants in our study. We saw similar tensions stemming from tactical non-disclosure manifest more dramatically in the following project.

Probing Community Values

This research was part of a larger effort to explore issues broadly related to sustainability [43] at a local farmers' market using cultural probes [9]. As part of this work, two of the authors sought to develop a methodology for doing "community probes," applying a similar approach and sensibility as cultural probes, but as a means of fostering conversations among community members rather than between community members and designers. Responses to a cultural-probe-style diary were used to inspire a series of such community probes in the form of speculative design proposals. For example, inspired by themes of stress and chaos in the diary responses, one design suggested a series of ropes and poles that automatically reconfigured themselves to allow for optimal foot traffic among the crowded market stalls. These proposals were intentionally provocative, as we sought to incite reactions from participants pertaining to what they liked and disliked about the market.

These designs were displayed on large posters at the market along with markers and post-it notes for visitors to leave comments. Adhering to the "aesthetic control" of cultural probes [18], the posters were not branded with the research lab or university where the authors worked. One of the central challenges that presented itself in the deployment of these cultural probes was that despite the satirical nature, or in some cases technical infeasibility, of the proposals, some market visitors thought not only that they were serious proposals for changes, but that these changes were being proposed by the market administration. When several market visitors complained about the proposals to the administration, we were asked to amend the posters with a disclaimer that the designs were in no way affiliated with or endorsed by the official market administration. We later followed up with the management, asking about conducting a follow-up where they could be more involved in the design process, but we were told that the market administration was not interested in running any more "surveys" at this time; in retrospect, these conversations illustrated not only a substantial misunderstanding of our research intervention, but also an antagonistic relationship between us and the public we wanted to engage.

Under some interpretations of cultural probes, we could read the overwhelmingly negative response to the community probe, both from the market attendees and the market administration, as a response and legitimate result of the probe: the refusal of even infeasible technologies points to community values around sustainability that oppose optimization narratives in favor of community gathering, revealing a community-oriented sustainability practice. We are very aware, however, of the interpersonal tensions that manifested when we came to the Farmer's Market and attempted to provoke critical responses from the public while temporarily masking our affiliations with the local university. While we were able to gather experiences and community responses to the probes we set up, the severed

relationship with the market administration points to the ways our relationships suffered as a consequence of the levels and types of information withheld.

THINGS GO AWRY

We initially found it difficult to articulate our experiences beyond a simple run-down of cases where we had “gone wrong” in doing critically oriented work, or a general indictment of empirical values unconsciously permeated into critically oriented HCI. Eventually, we realized that the evaluation practices researchers take to legitimate their research as working systems in HCI are deliberate, tactical engagements with the lingua franca, practices which in turn take a role in negotiating discursive norms in the community. For this reason, it is important not to treat our own tensions as isolated cases of research malpractice, but as actions situated in the context of the critically oriented community.

Using our own experience as a guide, we looked for traces of the same problematic in other critically oriented. We felt connections between the tensions in our work and those documented by Gaver et al. in “Anatomy of a Failure,” where the research team reflects on challenges in the evaluation of a sensor-based system, the Home Health Monitor, that was designed to be deliberately ambiguous in order to encourage interpretation and appropriation in domestic settings [20]. Similarly to our own methods, and in keeping with the studio’s practices, the researchers did not tell the participants in their study about their own intentions to avoid biasing the participants’ interpretations of the system. As the researchers describe, this initial lack of transparency contributed to a lingering attitude of uncertainty and suspicion that persisted even as researchers attempted to clarify their intentions through modifications in the system’s design. In the paper, the researchers discuss the ways in which withholding information and leaving the design open to interpretation made the system almost completely inscrutable and, simultaneously, almost completely uninteresting. We found points of resonance in their analysis in the ways that non-disclosure shaped the relationship between researchers and study participants.

Engagement between design researchers and their participants, and the ways that the relationship is made ever more complex by explicitly critical agendas, is also addressed in the work of Bardzell et al. [5]. In this work, the researchers deploy critical designs meant to provoke critical reflection on gender and divisions of domestic labor. This paper recognizes their participants’ struggle to participate (and be “good participants”) in the research. Like with Home Health Monitor, Bardzell et al. address aspects of participant suspicion through a desire to know what “the study is really about.” Part of this struggle in subject participation has to do with the idea that relationships between researchers and their participants do not start from a “blank slate.” One couple interviewed by Bardzell et. al. was particularly skeptical about the hidden motivations of the researchers stemming from their previous experiences in graduate school as

participants in psychological experiments. Bardzell et al. connect this struggle to the complexities of critical design, where the goal is to deal with uncomfortable topics (in their case, sex and gender) in deliberately provocative ways.

In our cases, we see traces of participant unease as to the nature of their relationship with researchers even in work that did not explicitly involve “critical design,” such as the speculative sketches we installed at the farmers’ market being (mis)interpreted as official proposals. This relationship becomes even more complex when the orientation of the participant toward the researcher is explicitly suspicious or even adversarial, as was the case with the activist communities in the Aoki et al. Street Sweepers fieldwork [4]. There, activists took on a generally oppositional stance toward academic researchers, who the activists saw as people who enter into activist communities for the benefit of their own research and disappear, along with the data, when this research is complete. Again, in navigating this complex relationship that comes as a unique consequence of employing subversive tactics in HCI, either in embodying alternatives in working systems or methods, we see noticeable frictions that arise when researchers negotiate different aspects of HCI’s lingua franca at different times in response to the specifics of their situation.

THREE DISCURSIVE TROPES

We have seen that in all the cases above, both our own and others from the literature, researchers are holding constant or appealing to some elements of more orthodox evaluation practices while challenging, bending, or exploring alternatives to others. How do we actually articulate what these practices are, to better understand what is being held constant and what is being changed? Here, we draw on values sourced from controlled scientific experimentation to highlight how HCI’s historical undercurrents of cognitive psychology and computer science shape elements of the current lingua franca in evaluation. For each of these three values, or discursive tropes, we describe how they can be employed by critically oriented researchers to creatively maneuver subversive research toward conventional notions of legibility. In practice, these three values are not clearly separable and are in fact bound up with one another in intricate ways. For the purposes of analysis, we tease them apart, but it should be noted that in practice they rarely separate cleanly. Our intention is to provide conceptual language with which to better understand both the nature and the ramifications of tactical engagement with discursive norms of evaluation in HCI.

Demand Characteristics

Psychologists and other social scientists have long acknowledged that the substrate they study—human beings—differs in important ways from that studied in the physical sciences. Namely, experimental subjects often reason, consciously or unconsciously, about the purpose of the experiment in which they are participating, not for

nefarious ends but often rather for compliance. Many experimental volunteers desire to be “good subjects” and see their participation as a contribution to the furthering of science [34]. The subject, then, also has a stake in the experiment turning out well. Thus, “as far as the subject is able, [s/]he will behave in an experimental context in a manner designed [...] to validate the experimental hypothesis” [34 p. 778, emphasis original]. Thus, an experimental result may not actually be due to the experimental manipulation itself but to the willing compliance of the subjects. Indeed, the subject may leverage a variety of cues – including the study description, the informed consent forms, the demeanor of the experimenter, the study procedures themselves - to reason (again, either consciously or unconsciously) about the purpose of the experiment. This “totality of cues which convey an experimental hypothesis to the subject [are called] demand characteristics” [34 p. 779].

Each of the above case studies demonstrates different attempts to reduce demand characteristics. For example, the work applying critical design to personal informatics was described to participants as dealing with personal data. We did not tell subjects that the design was “critical” in nature or that it was intended to prompt reflection on the value commitments that underlie personal informatics. In the frame reflection study, participants were told that the tool was about understanding political language, but framing was not explicitly mentioned. In the farmers’ market work, we did not want to tell participants that we were interested in sustainability so that they would not focus on a single aspect of their experiences at the market. This tactic followed from the advice of Gaver et al. [18] about ensuring that the speculative proposals had no trappings or accouterments of a university research lab, and other examples of researchers being deliberately ambiguous in the presentation of their design.

In each case, though, these decisions, intended rhetorically to minimize demand characteristics, ended up affecting, and in some cases limiting, the potential of our designs to engage participants in the very types of critical thinking and reflection we sought to engender. With the critical personal informatics study, participants expressed critical ideas about normative values in personal informatics in conjunction with literal interpretations of how the tool could be used to optimize behavior, because they suspected that this was our research goal. With the frame reflection system, participants were left a bit baffled as to how they might use the tool. Since frame reflection is admittedly a bit outside the normal approach to political news coverage, participants fell back on familiar forms of identifying and comparing partisan bias. At the farmers’ market, participants reacted most dramatically, thinking that these were official proposals rather than reflecting on broader themes about the market to which the speculative designs sought to draw attention.

We see maneuvers toward the reduction of demand characteristics across other critically oriented work in HCI, notably the techniques described earlier by Gaver et al. where researchers gather multiple perspectives on design work by informing neither the participants nor the independently hired film crew of their design intentions. Again, it is important to stress what while these negotiations appear ‘scientific’—and Gaver even compares the practices of evaluating Local Barometer to experimental evaluations done during his graduate work in Cognitive Science—the efforts to reduce demand characteristics are employed deliberately, under different (non-empirical) epistemic methodologies, to serve alternative ends [33 p. 140]. While this strategy can be rhetorically leveraged to tactically engage with HCI audiences, it can also create tension between researchers and study participants and impede researchers’ ability to pursue study goals.

Representative Sampling

Traditional evaluation techniques in HCI in their significant conceptual borrowing from empirical practices in experimental psychology and demographic sociology often center on questions of whether the participants in a study comprise a “representative sample”. Representative sampling is a statistical technique based on the idea that while it is rarely practical or even possible to get information from every person in a targeted population, the qualities of a population as a whole can be closely paralleled by a much smaller, properly selected segment. By selecting for some key variables such as gender, age, technical expertise, or socioeconomic status, and controlling for others, the representative sample becomes an effective ‘stand-in’ for other parts of that population. Traditionally in HCI, representative sampling emerges as a way to learn about potential users of a technology [38]. However, as HCI accrues interdisciplinary practices into its standard methodologies (e.g. interpretive and ethnographic ways of knowing,) the concern over representative sampling itself came to represent HCI’s internal tensions. As described in [33 p. 17], an ethnographer may refer to “sampling” a population to account and interpret their experiences and practices, and sometimes seek to make broader statements based on their interpretations without having their population “stand for” a broader statistical phenomenon in the same ways that controlled demographic or experimental approaches might attempt to do.

The question of representative sampling was salient at several moments of our research that are often not accounted for in traditional research accounts. For example, when writing on the critical personal informatics work went through peer review, one issue raised by reviewers was that our sample was not representative of the population as a whole, and that the results that we got might be specific to the participants we recruited. While we leveraged a rhetorical appeal to this same trope by arguing that our “sample” was representative of the audience to whom

personal informatics applications are traditionally marketed (i.e. upper middle-class, tech-savvy, college educated), in retrospect, this interaction was an example of the way that critically oriented researchers negotiate and enforce HCI's evaluation norms. Representative sampling negotiations are also visible in the Affecter research, where the authors point to concerns about "auto-biographical blinders" that might emerge from designers testing technology on themselves, and employed the use of a "third party" evaluator in response to this concern [42 p. 12:22].

As mentioned previously, the idea that research participants in critically oriented work can be used empirically to justify broader claims is also echoed by the very ways that we invoke scientific language of lingua franca to refer to participants in our writing (e. g. "P5" to refer to specific participants.) This is of course not to say that researchers literally dehumanize their participants into data, but rather that engaging with empirical epistemologies can be used rhetorically to make legible and render valuable to HCI some of the intellectual contributions of critically oriented researchers.

Stimulus-Response Causality

Experimental controls attempt to isolate various factors from one another in order to identify causal stimulus-response effects. To establish causal links, researchers often look for differences between an experimental condition and a control condition, where the two conditions are identical in every way with the exception of a single experimental manipulation: the stimulus. Thus, any differences observed between the two conditions must result directly from the stimulus. Generally, evaluations of critically oriented HCI are not situations of experimental control. Researchers rarely compare a control and experimental group or explicitly attempt to demonstrate causal relationships between some manipulated stimulus and some independent response. That said, many of the examples discussed above, both from our work and from others, evidence an interest in showing a type of causal relationship between a stimulus (i.e., design intervention) and a response (i.e., critical thinking or reflection).

For example, based on the project's motivation, we had a vested interest in demonstrating that the frame reflection tool itself, rather than any element of our study protocol, could effectively foster reflection on the framing of political issues. Similarly, we sought to show that the critical personal informatics themselves, and the generalizable strategies we used to design them, and not our interview questions, provoked people to consider the value commitments on which personal tracking technologies hinge (for example, to show that deliberate malfunction provoked critical reflection on infrastructure through reverse engineering.)

These evaluations were designed so that observed responses could be attributable to the system designs per se, rather

than to the evaluation protocol. It is relevant to note here that the effort to evaluate towards an imaginary where prototypes exist in the world and are used by people without the designers' intervention is not unique to critical projects in HCI, and that most (if not all!) HCI deployments employ elements of speculation. However, doing so tended to limit our ability to achieve some of the very goals our studies were intended to show we had accomplished, such as the double bind that occurred in the frame reflection example. Again, we see similar attempts to engage stimulus-response causality in others' work: the low-tech sensors research argues that the design of the sensors afforded particular types of usage. For example, by explicitly not attaching instructions about how the sensors worked, and then citing how all of the participants said the sensors were easy to use, the research team rhetorically convey that the devices were usable by traditional HCI standards. The research team was able to appeal to this trope of stimulus-response causality while also engaging participants in reflective and interpretive practices around environmental sensing, which illustrates the methodological creativity of critically oriented researchers to negotiate different values into the evaluation of their systems.

SYNTHESIS AND DISCUSSION

Broadly speaking, engagement with these tropes can legitimate new research directions while also shaping the nature of system evaluation and particularly the modes of engagement between researchers and participants. As we discuss in the next sections, these ramifications are not solely limited to individual research programs, and also bear important consequences for the critically oriented community as a whole.

Engagement and Tempered Radicals

In previous sections we described how, when compared to Agre's work in AI, the challenge of simultaneously engaging two audiences emerges as a quality unique to HCI. However, negotiating relationships between different communities while challenging the status quo is a practice that also finds parallels in other fields. We have found a helpful parallel between ideas espoused in critical technical practice and a concept from organizational science of the "Tempered Radical" [31]. Tempered radicals are individuals who identify with and belong to a certain organization while simultaneously being committed to a cause, community, and ideology that is fundamentally different from and at times at odds with the dominant values of the individual's organization. Tempered radicals seek to challenge the status quo by building up legitimacy within their organization and identifying strategic opportunities to negotiate change within their institution.

In spanning these boundaries, the tempered radical adopts an "ambivalent" stance that leads to tension. However, as Meyerson and Scully describe [31], the tempered radical can also be remarkably well positioned to assess where and

at which times “small wins” can be successfully enacted. Over the course of our analysis, we have been continuously impressed at the ways in which researchers have negotiated traditional HCI rhetoric and values to make room for and legitimate new research programs. Though this work is not without risks of “loss” or re-absorption into mainstream practices (e.g., in the case of [11]), we feel that critically oriented research practice can be used, gainfully, to challenge status quo values and practices in the design and evaluation of technology.

Politics of Evaluation

An essential component for the success of the “tempered radical” is their affiliation with other like-minded members of their organization. Previously, we have described critically oriented researchers as people engaging simultaneously with two audiences, their participants and the broader HCI academic community. However, methodological decisions made by critically oriented researchers also impact a *third* group of people, namely other critically oriented researchers in HCI. Here we channel an argument made by Cohn et al. that methods used in designing and evaluating have more general consequences for discursive and practical action because methods enable specific discourse and forms of knowledge production [14]. In other words, just as the HCI community enforces a set of discursive norms, so do critically oriented sub-communities within HCI. In this sense, critically oriented researchers should be conscious of their role in normalizing practices for future researchers who wish to engage in related forms of subversion. We recognize that mainstream HCI publications may not always be the most appropriate venues for these discussions (although it has certainly been possible to write about tensions while couching them in lingua franca rhetoric of success, e.g. [20]). We do believe that conversations among critically oriented researchers lead to helpful research contributions (e.g. [20, 5, 32]) and hope that our analysis can be used to articulate the trade-offs that critically oriented researchers make when engaging with HCI practices.

CONCLUSION

By looking back on our research and by thinking through our analysis of the values we engage and their trade-offs, it becomes possible to imagine potential future trajectories for our projects. For instance, in the frame reflection example, we could change the conditions of the demand characteristic reduction to allow sharing information about framing that we previously withheld from participants. Instead, we could re-negotiate the “response” portion of the stimulus response to ask not whether the system makes people reflect on framing but rather to evaluate the ways in which participants then conceptually engage with ideas of framing using the system. More broadly, what we hope to contribute through this analysis is not a generalizable method or framework for critically oriented evaluation or engagement with the traditional HCI lingua franca. Rather,

we offer a generalizable stance or lens that critically oriented researchers can use to articulate their own tactical engagements with HCI’s discursive norms. We believe that such conversations will help critically oriented researchers work with each other to acknowledge and innovate evaluation methods in HCI.

ACKNOWLEDGEMENTS

We would like to thank the NSF under Grants Nos. IIS-1110932 and IIS-1217685; the Intel Science & Technology Center for Social Computing; Melissa Mazmanian and the Doctoral Research and Writing Seminar at UC Irvine; and the anonymous reviewers for their wonderful feedback.

REFERENCES

1. Agre, P. *Computation and human Experience*. Cambridge University Press, 1997.
2. Agre, P., Chapman, D. 1987. Pengi: an implementation of a theory of activity. In Proceedings of the sixth National conference on Artificial intelligence - Volume 1 (AAAI'87), Vol. 1. AAAI Press 268-272.
3. Agre, P. "Toward a critical technical practice: Lessons learned in trying to reform AI." *Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work*, Mahwah, NJ: Erlbaum (1997): 131-157.
4. Aoki, P. M., Honicky, R. J., Mainwaring, A., Myers, C., Paulos, E., Subramanian, S., & Woodruff, A. "A vehicle for research: using street sweepers to explore the landscape of environmental community action." *In Proc. CHI '09*
5. Bardzell, S., Bardzell, J., Forlizzi, J., Zimmerman, J., and Antanitis, J. "Critical design and critical theory: the challenge of designing for provocation." *In Proc. DIS '12*
6. Bardzell, J., Bardzell, S, and Stolterman, E. 2014. Reading critical designs: supporting reasoned interpretations of critical design. *In Proc. CHI '14*
7. Bardzell, J., and Bardzell, S. 2013. What is "critical" about critical design?. *In Proc. CHI '13*
8. Baumer, E. P. S., Cipriani, C., Davis, M., He, G., Kang, J., Jeffrey-Wilensky, J., Lee, J., Zupnick, J., and Gay, G. K. (2014). Broadening Exposure, Questioning Opinions, and Reading Patterns with Reflex: a Computational Support for Frame Reflection. *Journal of Information Technology & Politics*, 11(1), 45–63
9. Baumer, E. P. S., Halpern, M., Khovanskaya, V., & Gay, G. Probing the Market: Using Cultural Probes to Inform Design for Sustainable Food Practices at a Farmers' Market. In J. H. Choi, M. Foth & G. Hearn (Eds.), *Eat, Cook, Grow: Human-Computer Interaction with Human- Food Interaction*. MIT Press, 2014.

10. Blythe, M., Overbeeke, K., Monk, A., Wright, P. (eds). *Funology: From Usability to Enjoyment*. Kluwer Academic Publishers, 2003.
11. Boehner, K., Vertesi, J., Sengers, P., & Dourish, P. "How HCI interprets the probes." In *Proc. CHI '07*
12. Boehner, K., Sengers, P., and Warner, S. 2008. Interfaces with the ineffable: Meeting aesthetic experience on its own terms. *ACM Interact.* 15, 3, Article 12 (December 2008), 29 pages.
13. Chong, D., & Druckman, J. N. (2007). Framing Theory. *Annual Review of Political Science*, 10(1), 103–126.
14. Cohn, M., Sim, S., and Dourish, P. 2010. Design methods as discourse on practice. In *Proc. GROUP '10*
15. DiSalvo, C. *Adversarial Design*. The MIT Press, 2012.
16. Dourish, P. *Where the action is: the foundations of embodied interaction*. MIT press, 2004.
17. Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4), 51–58
18. Gaver, W. W., Boucher, A., Pennington, S., & Walker, B. (2004). Cultural probes and the value of uncertainty. *interactions*, 11(5), 53-56.
19. Gaver, William W., Jacob Beaver, and Steve Benford. "Ambiguity as a resource for design." In *Proc. CHI '03*.
20. Gaver, W., Bowers, J., Kerridge, T., Boucher, A., & Jarvis, N. "Anatomy of a failure: how we knew when our design went wrong, and what we learned from it." In *Proc. CHI '09*.
21. Gaver, W., Boucher, A., Law, A., Pennington, S., Bowers, J., Beaver, J., Humble, J., Kerridge, T., Villar, N., and Wilkie, A. "Threshold devices: looking out from the home". In *Proc. CHI '08*
22. Goffman, E. (1974). *Frame Analysis*. Cambridge, MA: Harvard University Press.
23. Hirsch, T. 2010. Water wars: designing a civic game about water scarcity. In *Proc. DIS '10*
24. Khovanskaya, V., Baumer, E.P.S., Cosley, D., Vaida, S. & Gay, G. "Everybody knows what you're doing: A critical design approach to personal informatics." In *Proc. CHI '13*
25. Kuznetsov, S., Hudson, S., and Paulos, E. "A low-tech sensing system for particulate pollution." In *Proc. TEI '14*
26. Kuznetsov, S., Davis, G. N., Paulos, E., Gross, M. D., & Cheung, J. C. 2011. Red balloon, green balloon, sensors in the sky. In *Proc. UBICOMP '11*
27. Kuznetsov, S., Davis, G., Cheung, J., & Paulos, E. "Ceci n'est pas une pipe bombe: authoring urban landscapes with air quality sensors." In *Proc. CHI '11*
28. Kuznetsov, S., Odom, W., Moulder, V., DiSalvo, C., Hirsch, T., Wakkary, R., & Paulos, E. HCI, politics and the city: engaging with urban grassroots movements for reflection and action. In *CHI EA '11*.
29. Lieberman, H. The Tyranny of Evaluation, <http://web.media.mit.edu/~lieber/Misc/Tyranny-Evaluation.html>
30. Löwgren, J. 2013. Annotated portfolios and other forms of intermediate-level knowledge. *interactions* 20, 1 (January 2013), 30-34.
31. Meyerson, D., and Scully, M. "Crossroads tempered radicalism and the politics of ambivalence and change." *Organization Science* 6.5 (1995): 585-600.
32. Michael, M. (2012). "What are we busy doing?" Engaging the idiot. *Science, Technology & Human Values*, 37(5), 528-554.
33. Olson, J., and Kellogg, W. *Ways of Knowing in HCI*. Springer, New York, NY, 2014.
34. Orne, M. T. (1962). On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776–783.
35. Pierce, J. 2012. Undesigning technology: considering the negation of design by design. In *Proc. CHI '12*
36. Pierce, J. and Paulos, E. "Counterfunctional things: exploring possibilities in designing digital limitations." In *Proc. DIS '14*
37. Polletta, F., Pierski, N., Baumer, E. P. S., Celaya, C., & Gay, G. (2014). A "Peopled" Strategy of Frame Reflection. In Annual Meeting of the American Sociological Association (ASA). San Francisco.
38. Preece, J., and D. Maloney-Krichmar. "The human-computer interaction handbook." (2003): 596-620.
39. Schön, D. A., & Rein, M. (1994). *Frame Reflection: Toward the Resolution of Intractable Policy Controversies*. New York: Basic Books.
40. Sengers, P., Boehner, K., Mateas M., and Geri Gay. 2008. The disenchantment of affect. *Personal Ubiquitous Comput.* 12, 5 (June 2008), 347-358.
41. Sengers, P., Boehner, K., David, S., & Kaye, J. J. "Reflective design." In *Proc. CC '05*
42. Sengers, P., and Gaver, B. "Staying open to interpretation: engaging multiple meanings in design and evaluation." In *Proc. DIS '06*
43. Silberman, M., Blevins, E., Huang, E., Nardi, B. A., Nathan, L. P., Busse, D. Preist, C., and Mann, S. "What have we learned?: a SIGCHI HCI & sustainability community workshop." In *CHI EA '14*.
44. Suchman, L. *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press, 200