

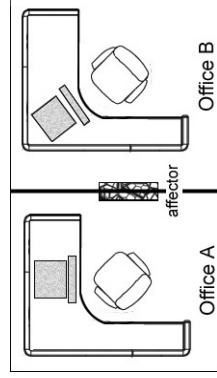
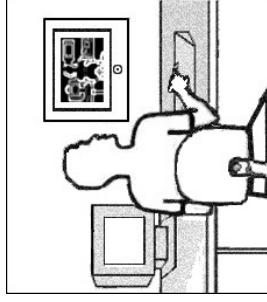
Closing the Affective Gap & Opening Up Evaluation

Kirsten Boehner

Information Science Building
301 College Avenue
Ithaca, NY 14850
kab18@cornell.edu

Phoebe Sengers

Information Science Building
301 College Avenue
Ithaca, NY 14850
sengers@cs.cornell.edu



Figures 1 and 2: Sketch of Affecter Set Up

Introduction

One driving objective in affective technology is that by incorporating affect our designs will result in richer, and more natural, experiences. Unfortunately, what often results with this new affective technology tends to have the opposite result of flat, reduced representations. We argue this is due to an under acknowledged gap, the *affective gap*, between human emotional experience and its formalization. One popular approach to close the affective gap is to improve computers' formalized models to address the full complexity of human emotional experience. In this approach, the gap is closed through efforts to design computers with 'human-like' abilities regarding emotion. As an alternate approach, we propose to close the gap by rebalancing the roles of humans and machines in affective computing. In our model, affect is co-constructed through a process of collaboration. Humans do what they are good at - situated interpretation and experience of emotions - while computers support these activities with what they are good at - e.g., data gathering and storage, real-time communication, and formalized representations of human behavior.

Together with our collaborators, led by Kristina Höök at SICS, Bill Gaver at Goldsmiths, Geri Gay at Cornell, and Michael Mateas at the Georgia Institute of Technology, we explore affective presence systems that support reflection on rich, enigmatic experiences of affect. Our approach is informed by four underlying principles that influence not only how we design for affective presence but also what it means to evaluate such systems. These principles are positioned as shifts from traditional work in HCI in general and affective computing in particular that bias formalized systems:

1. We shift from affect as information, or discrete transferable data, to affect as interaction ([1],[13])
2. We shift from classifying and responding to affective data to supporting rich co-interpretation of affect between people and machines ([11],[14]).
3. We shift inspiration from perspectives and methods based in cognitive psychology to a broader set from the arts and humanities ([6],[7],[15]).
4. We shift from a primary focus on what computers can know about emotion to how people can become aware of and reflect on their own and each others' emotions ([9],[12]).

A Case Study of Co-Interpretation

To illustrate the above principles and subsequent issues of evaluation, we will make a case study of Affecter (Fig. 1 & 2), a digital window between friends in neighbouring offices that allows glimpses of one another's moods [10]. Mood is communicated by systematically distorting a video feed between the offices, thereby attempting to both protect privacy and enhance the emotional ambiguity of the display. Following Brooks's principle of intelligence without representation [2] there is no representation of emotion within the system; instead, machine vision and sound data directly trigger distortions, which are read for their emotional meaning by users in the context of their relationship and interactions. The display is meant as a demonstration of affect rather than a symbolic representation of affect. Disambiguating system output is in the province of the systems' users, drawing on their existing rich understanding of one another based on their day-to-day interaction.

The four principles described in the introduction directly influence what Affecter is designed for and the design strategies employed -- most notably in this case the use of autobiographical design, an approach that preferences idiosyncratic and personal experiences [8]. For this paper, however, we will focus on the evaluation methods selected to resonate with our design principles and strategies.

Approaching Evaluation

As the theoretical principles of affect influence the goals and strategy of design for affective presence, we shift from evaluating the accuracy of the representation to the engagement and evocativeness of the experience. Our questions then shift from "did the system get it

right?" to "how is it that the users define a 'working' system?" Since our design methods are inspired by theoretical approaches to meaning making and interpretive acts, our evaluation goals and methods also draw from forms of interpretivist and critical inquiry [3]. Using Affecter as our case study, we will outline below four areas of evaluation focus driven from our previous principles.

1. ENGAGING USERS IN A PROCESS OF MEANING-MAKING

In order to assess this process, we are informed by qualitative evaluation methods such as *account analysis* through *observations* of the system in use and *open-ended interviews*. With Affecter, we observe how the designers shift into use mode when the system is running in their offices. We watch to see how they orient to the display screen. Do they incorporate the screen into their office surroundings and practices? Occasionally, an evaluator will enter one of the offices and ask the user to narrate what is happening with the system at that particular moment and how he or she used the system that day, if at all. We also periodically ask the users, this time as designers, to tell us what the goals of the system are and how they see these changing given their experiences and new expectations.

In addition to examining how a given designed system is appropriated, we explore the effects of systematic variation of attributes and contexts. For example, we vary how stringently we hold to Brooks' maxim of no internal representation to see how a stronger system representation influences how the users/designers narrate the 'working' system. These system attributes will be manipulated under alternate conditions, for example absence, randomness, and informatics. For the absence condition, we look at the changes pre and

post a system's implementation. For the randomness condition, we substitute a random generator rather than selected distortions by the users/designers to assess if the system is doing more than acting as a simple screen saver or a lava lamp. Finally, for the informatics condition, this is a test of the same system but implemented for affect as 'information' versus affect as 'interaction.' For example, in Affecter, we test its use against a video conferencing system – one that prioritizes straight information transfer. We may also test Affecter with an internal map of emotions and a display that presents this as an information graph, for example "Simeon's level of happiness today is 5."

2. ACCOUNTING FOR SHARED OR MULTIPLE INTERPRETATIONS

As described in the above section, we use qualitative methods to develop a phenomenological account of how the users/designers engage with the system to develop an awareness of or express affect. But we are not interested in a single users/designers interpretation but how interpretation is coordinated between designers, users, and system. To this end, we use *inter-account analysis* and pay particular attention to how a user communicates his or her interpretation and how this interpretation is then taken up or challenged by others. Informed by conversation analysis, we look at what evidence is presented by each party in terms of uptake and understanding and how this evidence is either agreed with, repaired, or perhaps ignored. In the Affecter system, we examine how the users talk about the system in use and in reflection. We are interested in ascertaining if they talk about the system in different ways when they are wearing the hat of designer versus when they are wearing the hat of user. We look to not only describe these different accounts but to examine

how, if at all, these accounts rub up against and transform each other.

3. EVOKING RICH DETAILED NARRATIVES OF ENGAGEMENT

In our evaluations, we seek to provide users with the opportunity to share the richness and details of their evoked experience. One tool we have used for this end effectively is the use of *guided journals* [5]. The Affecter system's configuration interface comes with a logging feature allowing the users/designers to detail why certain changes are made to the system. In future iterations of this interface, we will include the ability to insert screen shots in order to provide visual prompts or illustrations for the narrative that unfolds. We also plan to draw from the design method of *cultural probes* [4] to prompt and inspire reflection by the users/designers. The guiding objective here is to make the reflection process itself an engaging and enjoyable experience for the users, something far different from the task of circling numbers on a seven-point scale.

4. STIMULATING REFLECTION ON AFFECT AND TECHNOLOGY

The last principle of affective systems is to promote reflection not on the technology by itself but on expressions of affect, the nature of affect, and the role of technology in this construction and expression. In order to assess how affective presence systems accomplish this, we document how, when, and what kind of reflections occur. This is partly assisted through the use of tools to evoke personal narratives as described above. In this same spirit, we design evaluation metrics into the systems as a form of *dynamic feedback*. In this way, our evaluation methods are construed as part of the designed system, not something that occurs in parallel or subsequent to the design and eventual use.

One form of dynamic feedback information is usage data. In a traditional HCI evaluation, this information would be tracked as a metric of success for the evaluator to analyze. In affective presence systems, we disclose this *data as a stimulus* for users to narrate what is happening. For instance, with Affecter, the users/designers log how often they interact with each other and how often they look at Affecter. Fewer interactions may signal that Affecter supplies emotional interconnection that was previously missing, or that it makes users want to interact less. Interpreting the significance or meaning of this data is not left to the evaluators of the system but is interpreted and incorporated as a resource by the users/designers.

Conclusion

In moving to the principles of affective presence systems, we open up the question of evaluation, as traditional HCI methods must be appropriated and new methods created. In fact, many of the traditional HCI methods are violated in our evaluation strategies. Just as the line between users and designers is merged in the Affecter system, the line separating evaluation also becomes more permeable. We do use outside evaluators, but their interpretation of what is going on is not privileged above that of the users/designers. This is evident in our evaluation methods, the journals for example must not only enact for the evaluators what the experience feels like or means, but they provide a resource for the users as well to reflect on and perhaps influence this experience.

In traditional HCI and affective computing studies, self-report methods are viewed with a degree of suspicion: the user could be misrepresenting the truth either because he or she forgot certain details or the

experience became transformed in memory or perhaps the user wants to present a more idealized self to the evaluators. Evaluations of affective presence do not deny that these phenomena may be at play but we are interested in what the system comes to mean to the users – whether this be based on some ‘truth’ or if the ‘truth’ is called into being. In other words, the very reasons why self-report methods are problematic in traditional HCI evaluation are reasons of interest for affective presence evaluation.

As another example of divergence, a traditional HCI evaluation might not present usage statistics and patterns back to users for fear that this information might subsequently influence how the users interact with the system or color their perceptions of the systems’ performance. Again, for our purposes, this kind of reaction would constitute an interesting result as opposed to a flaw in the evaluation design.

We propose that the principles and design strategies of affective presence open up evaluation to new forms, new objectives, and new criteria for judging the evaluation itself. Three particular instances detailed here include the merging between the designed system and the evaluation design. It is difficult in our systems to draw out where the design ends and the evaluation begins. Secondly, we underscore the importance of evaluating for particularities, for the idiosyncrasy of the individual and the situation. These particularities, rather than abstract commonalities, become the basis for generalizing lessons across different systems. Finally, we look to foster multiple interpretations rather than single verdicts on success or failure. We believe this opening up of evaluation will play not a summative but a formative role in the closing of the affective gap.

Acknowledgements

This work was funded in part by a National Science Foundation grant 0238132. Affecter is a collaboration with Simeon Warner, Eunyong "Elie" Shin, David Klein, Rev Guron, Eugene Medynskiy, and Liz Goulding. We would also like to thank our many colleagues who inspire this work: Joseph 'Jofish' Kaye, Geri Gay, Kristina Höök, Petra Sundström, Jarmo Laaksohlahti, Bill Gaver, Michael Mateas, Paul Dourish, Rogerio De Paulo, Katherine Isbister, Peter Wright, John McCarthy, and Mark Blythe.

Works cited

- [1] Boehner, K., R. DePaula, P. Dourish, and P. Sengers. Affect: From Information to Interaction. Proceedings of the Critical Computing Fourth Decennial Aarhus Conference. Aarhus, Denmark, (2005), 59-68.
- [2] Brooks, R. A. Intelligence without representation. *Artificial Intelligence*, 47(1-3), (1991), 139-159.
- [3] Denzin, N. and Y. Lincoln. Handbook of Qualitative Research, 2nd edition. Thousand Oaks, CA: Sage Publications, (2000).
- [4] Gaver, W., Dunne, A., & E. Pacenti. Cultural Probes. *Interactions* 6(1), (1999), 21-29.
- [5] Kaye, J., Levitt, M., Nevins, J., Golden, J., & V. Schmitt Communicating Intimacy One Bit at a Time. *CHI'05 Extended Abstracts*, ACM Press, (2005), 1529-1532.
- [6] Höök, K., Sengers, P. & G. Andersson.. Sense and Sensibility: Evaluation and Interactive Art. In *Proceedings of the 2003 Conference on Human Factors in Computing Systems*, Ft. Lauderdale, Florida, (2003).
- [7] Mateas, M. Expressive AI: A hybrid art and science practice, In *Leonardo: Journal of the International Society for Arts, Sciences, and Technology*, 34 (2), (2001), 147-153.
- [8] Sengers, P. Autobiographical Design. *Digital Arts and Culture 2005 Workshop on The Theory and Practice of Experience Design*, (2005).
- [9] Sengers, P., Boehner, K., David, S., and J. Kaye. Reflective Design. Critical Computing Fourth Decennial Aarhus Conference. Aarhus, Denmark, (2005), 49-58.
- [10] Sengers, P., Boehner, K., Warner, S., and T. Jenkins. Evaluating Affecter: Co-Interpreting What 'Works'. *CHI 2005 Workshop on Innovative Approaches to Evaluating Affective Systems*, (2005).
- [11] Sengers, P. & B. Gaver. Designing for Interpretation. *Proceedings of Human-Computer Interaction International*, Las Vegas, Nevada, (2005).
- [12] Sengers, P., Liesendahl, R., Magar, W., Seibert, C., Müller, B., Joachims, T., Geng, W., Mårtensson, P., & K. Höök. The Enigmatics of Affect. In *2002 Conference on Designing Interactive Systems*, London, England, (2002).
- [13] Suchman, L. Figuring Personhood in Sciences of the Artificial. Department of Sociology, Lancaster University, (2004).
- [14] Sundström, P., Ståhl, A., & K. Höök. eMoto – Affectively Involving both Body and Mind. *Proceedings of the 2005 Conference on Human Factors in Computing Systems*, Portland Oregon, (2005).
- [15] Wright, P. & J. Finlay, J. arts-hci.org/lit-hci.org *Workshop on HCI, the Arts and Humanities*. University of York, York, England, (2003)