

Sense and Sensibility: Evaluation and Interactive Art

Kristina Höök¹, Phoebe Sengers², and Gerd Andersson¹

¹SICS
Box 1263,
164 29 Kista, Sweden
+46 8 633 15 00
{kia, gerd}@sics.se

²Cornell Information Science and S&TS
301 College Ave.
Ithaca, NY 14850 USA
+1 607 254-5396
sengers@cs.cornell.edu

ABSTRACT

HCI evaluation methods are useful for improving the design of interactive systems, yet they may be rejected by nontraditional technology disciplines such as media art. We have developed a two-tiered evaluation model that responds to the concerns of interactive artists and have used it to improve the design of an interactive artwork, the Influencing Machine, exploring issues in affective computing. The method was interpretive, focusing on giving the artists a grounded feeling for how the machine was interpreted and their message was communicated. We describe the resulting design of the Influencing Machine and the reactions of users. The study itself is part of the art piece – together these activities achieve the goal of the artists: to provoke our cultural notions of whether a machine can “have emotions”.

Keywords

Evaluation, Affective Computing, Interactive Art

MOTIVATION

There is a conflicted convergence developing between human-computer interaction and interactive art. Artists are building interactive computational systems, such as Natalie Jeremijenko’s Dangling String [15], that develop fresh new perspectives on interactive system design. HCI practitioners are adapting ideas from art practice to rethink fundamental problems in HCI [e.g. 5]. But art and HCI are not easily combined. While HCI researchers develop new technologies, methods, and standards, interactive artists adapt, alter, and challenge them with very different goals in mind. In this hybrid area, fundamental conflicts in worldview and methodology arise that can propel both fields forward but need to be negotiated carefully.

In particular, HCI strategies for evaluation have been proven an effective tool to understand and improve digital

systems. Yet interactive art generally ignores HCI methodologies for evaluation, based on a mostly unstated belief that they do not measure aspects of interactive artworks that are of interest to artists. In this paper, we analyze the mismatches between standard evaluation methodology and the perspectives of artists, and develop a new evaluation methodology that may be more appropriate to the concerns of artists. We use this methodology to evaluate an interactive artwork, the Influencing Machine, which explores the relationship between users and affective computers.

It would be ludicrous for us to suggest replacing art criticism with HCI evaluation, and we will not answer the question “is this good art?” But we will show that, suitably adapted, user testing can help fine-tune the interaction design of interactive artwork, helping artists to get their message across. In the process, we also hope to show how the perspective of artists can help HCI evaluation by suggesting some new aspects of the relationships between system builders, users, and evaluators. The aims of this paper are both to show the specifics of the Influencing Machine and how its users understood it, and also to inspire a better understanding and relationship between artists and HCI evaluation methods.

USER STUDIES AND INTERACTIVE ARTWORK

Within HCI, formal user studies (ethnographic or quantitative-scientific) are the gold standard for evaluating computational systems. Artwork, in contrast, is generally evaluated using the tools of art criticism, in which non-artist experts develop a rigorous, subjective argument about the artwork by analyzing it with respect to broader trends in the art world. Formal ‘user’ studies play no role. For artists, users are not an object of study but themselves the final target of their work. An interactive artwork is, in this sense, less an object to be tested and more like a research paper: the artist uses it to communicate his or her ideas directly. Hence, laboratory evaluation of artwork can seem as ridiculous to artists as laboratory evaluation of CHI papers would to an HCI researcher.

As a consequence, hybrid artwork-interfaces are often evaluated through a combination of argument and informal conversation with users (e.g. [10, 13]). Similar informality can be observed in the HCI literature on evaluation of art-influenced speculative design. For example, the Presence

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2003, April 5–10, 2003, Ft. Lauderdale, Florida, USA.
Copyright 2003 ACM 1-58113-630-7/03/0004...\$5.00.

project was evaluated informally by describing the designers' experience in installing the system and observing user interaction [5].

A more formal evaluation was done by the Placebo project, involving art-like furniture that responds in different ways to electromagnetic fields in the home [3]. This study differed from a traditional HCI study in a number of ways. Users were recruited for the study using a form that would influence their experience of the project: they were asked if they would like to 'adopt' furniture and if they had had unusual experiences with electromagnetic fields. Instead of a large laboratory study, a small number of users were given the objects for a period of time to have in their homes. Designers came to users' homes to interview them. The results of the study are the raw texts of the interviews, not further analyzed in any way. In addition, a portrait photographer took art-style photographs of the subjects with their devices. These photos are presented as part of the evaluation, again without further analysis or interpretation. The lack of analysis leaves the reader of the study free to draw their own conclusions – it is open to interpretation. Nothing is proven, no arguments are made, and there is no simple list of conclusions.

Anecdotal evidence, informal chats between users and system-builders, tiny study sizes, forms structured to influence user interpretation, no discussion or analysis of results: this may sound like a to-do list for bad evaluation. But these choices are deliberately and thoughtfully made, highlighting underlying conceptual problems in using standard HCI techniques to evaluate art-inspired systems. Grossly speaking, the major conflict between artistic and HCI perspectives on user interaction is that art is inherently subjective, while HCI evaluation, with a science and engineering inheritance, has traditionally strived to be objective. While HCI evaluation is often approached as an impersonal and rigorous test of the effects of a device, artists tend to think of their systems as a medium through which they can express their ideas to the user [10] and provoke them to think and behave in new ways. When artists do use user studies, they are likely to see the user study itself as part of the communication through the artwork and another opportunity to shape the 'message' of the artwork. This can be seen clearly in Garabet et al.'s use of performance art as a provocative kind of user study [4].

The subjective approach means artists do not build systems for 'normal' or 'average' users. Artists are interested in the richness and complexity of unique, individual users, cultural contexts, and resulting variety of interpretations and experiences of their system. Artists see not only users, but also the readers of user studies, as engaging in complex acts of interpretation; hence it is not appropriate to summarize the results of a study into a few statements that are said to hold for everyone. Also, the statistical averaging and laboratory simplifications necessary for reliable scientific statements may wash out all the details that interest them.

Artists may prefer a rich, narrative, and singular understanding to a simpler but rigorous and generalizable understanding. This interest in singularity and narrative complexity allies well with the recent ethnographic turn in HCI; yet many ethnographers may feel uncomfortable in promulgating a personal vision to users to the same extent as artists.

These mismatches leave us with two questions. First, what role could or should user studies play in the evaluation and development of interactive art? Second, how should user testing strategies be altered to be appropriate to the concerns of artists? We develop one approach and use it to improve and understand the Influencing Machine, an interactive artwork by Phoebe Sengers, Rainer Liesendahl, Werner Magar, and Christoph Seibert.

CASE STUDY: THE INFLUENCING MACHINE

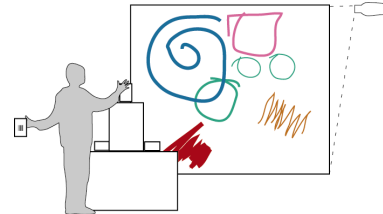


Figure 1. Setup of the Influencing Machine

Two people enter a small room. Child-like scribbling appears across a wall: jagged lines, circles, spirals, and other shapes build up, overlap, fade away (see figures 1 and 3). Scattered throughout the room are postcards with art prints or color fields; on a table stands a wooden mailbox (see figure 2). One person picks up a card and tentatively puts it in the box. Unusual and musical sounds begin to play. Drawings change speed, color, pressure, form. The people begin sorting through cards, dropping them in the box and seeing how the graphics and sound change. They play, experiment, discuss: "How is this reacting to us?" "How do you think this works?"



Figure 2. The constructed Mailbox.

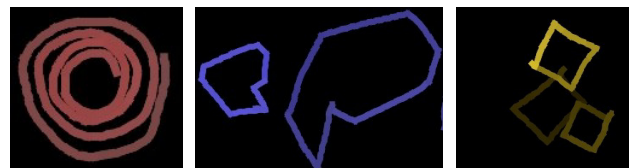


Figure 3. Examples of generated scribbles.

Technically, the system works by using the input postcards marked with machine-readable bar codes to influence an internal emotional model. These internal emotions trigger

sounds and the selection of drawing behaviors and their dynamic parameters: speed, color, size, pressure, etc. When the machine receives input, system drawings tend to become gradually more complex; when it has not received input for several minutes it restarts. While this technical description is precise and clean, the emotional interpretation of the graphical output and postcards by users is complex, incompletely specifiable, open-ended, and strongly culturally influenced. The Influencing Machine explores the tension between machines and affective beings in affective computing; how will people relate to a machine whose emotions they can influence, but whose behavior they cannot control [14]?

EVALUATION STRATEGY

But what exactly were we going to check once we brought the Influencing Machine and users into the lab?

The purpose of the Influencing Machine is to create a cultural provocation, challenging our views of what a machine can be, in particular whether it was capable of being emotional – but how would we check what the machine in fact was able to provoke? What if users did not get the idea at all, or if they only got frustrated and dismissed it entirely? A provocation entails an experience that is not necessarily easy or pleasant for users, so we may have the goal of developing painful or difficult situations. This is something standard usability strategies will try to avoid.

We had to disentangle frustration that came from *bad design* choices from frustration that came from actually encountering a machine that cannot be controlled – only influenced. The design of the Influencing Machine is balancing on a thin line between being predictable and controllable and thereby boring and not achieving its purpose, and being unpredictable and uncontrollable and thereby alienating its users, making them feel stupid and out of control entirely.

Since this evaluation would happen during the early design cycles of the machine, it would seem as if normal usability testing in the lab was the only possibility. Methods based on activity theory [11] or contextual design [2], that seek to understand the entire complexity of a situation, including culture, tools, and practices, require that the new tool is made part of the culture for a longer period, and that it can be studied in its context for a longer time span, integrated with the culture. An artist, on the other hand, often wants to challenge the current cultural notions. The Influencing Machine, presented here, is a provocation of current mainstream ideas of what computers are. Thus studying it in its context for a longer time is not what we want.

Thus, for many reasons, we were willing to find a compromise between a natural, contextual evaluation in the full complexity of real life and an unnatural (but as natural as possible) lab situation where subjectivity is central. It should be noted that our contribution here is not an entirely new method for interactive design of interactive art – instead we believe that the art piece itself, the evaluation of

it, and this paper describing the two, go together in achieving the goal of the artist: to provoke our thinking on affective machines. We also believe something can be learned about the specific difficulties of understanding how to evaluate and help redesign interactive artworks.

Methodology

We decided to place the Influencing Machine in the lab to get feedback about how to (possibly) redesign it so that the intentions behind the machine are reasonably conveyed. Putting it in the lab made us run into several practical method issues.

Subjects in groups

We wanted to know more about users' subjective understandings of the machine, their theories of what it was, their metaphors for how they described it, but we did not want to disturb their experience through forcing them to speak aloud. We also wanted to distract them from being too aware of the laboratory situation. We decided to use the co-discovery method [2] where users are brought in two and two, with some slight modifications. We brought in users in different group sizes. Also, we were not interested only in the talk-aloud effect, but also in group dynamics around the art piece. We would not get any reliable "average user" data, but, on the one hand, this is closer to how art is often experienced and, on the other, group reactions and dynamics are more interesting than average, single, normal users. Since subjects spoke to one another naturally we could follow their theory-forming process – as well as be given insights into how their personalities interacted with the machine.

Users backgrounds

As we were bringing subjects into the lab, we also needed to decide on *who* to bring in – something an artist only indirectly can decide through placing his/her piece of art in selected surroundings.

Our previous experiences from evaluating a system named Agneta & Frida [8] made us see that it is not enough to classify users by age, gender and education. In order to pinpoint finer distinctions in users' emotional reactions, we have to consider users' interpretation, understanding, attitudes, personality and expectations of computer culture.

Narrative context, metaphors used

The narrative context of the Influencing Machine is made purposefully unclear. In addition to recording the subjects' interactions with one another, open-ended interviews were done to better understand the metaphors users use to describe the machine [8, 9]. Metaphors are interesting because they help us to think about relatively abstract conceptual domains in terms of relatively concrete domains. The question here is what users will say about the Influencing Machine? Will they talk about it as a child, a piece of art, an agent, or a computer application?

Setting in the laboratory

The room in which the Influencing Machine was placed did not look like an office or like a laboratory (see figure 4.). Instead the room was in an old house, it had a Dutch tile stove, arts on the walls, and shelves with books. There was a high table on which the machine was placed, rather than an office desk. There were no chairs around the machine; postcards were scattered around the room. The computer was hidden under a table with tablecloth on it.

Evaluation in two steps

Affective interaction systems must, similar to intelligent user interfaces in general, be evaluated in two steps [7]. First, we must make sure that the users understand the emotions expressed by the system, or that the emotions expressed by users are understood by the system. Second, we need to check whether this in fact leads to the desired effects on the overall interaction. It might be that the overall design of the Influencing Machine is perfectly valid, but the drawings are hard to interpret as emotions. Or conversely, the emotional expressions might be easily understood by the user, but the design does not achieve its overall goal of provoking or charming the user.

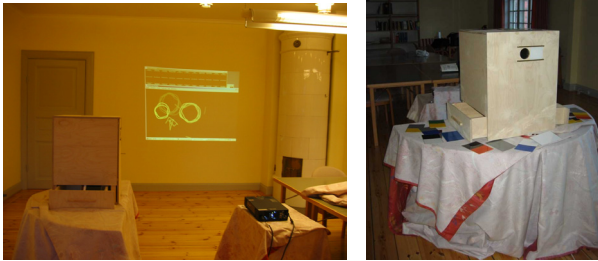


Figure 4. The setting in the lab.

Method

The subjects were first interviewed about their attitude towards art installations, whether they had children, and other demographic information.

They were told that the machine was “about emotions and that they would be posting postcards into a mailbox”. They were encouraged to interact as they pleased with the machine and to stop whenever they got bored. We informed them that they would be videotaped.

Afterwards we asked a set of questions in an open interview. These questions are carefully designed not to privilege a particular metaphor of the Influencing Machine – not to talk about the machine as a machine, nor as a s/he.

Subjects

The first study had six groups with a total of 12 subjects, aged between 20 and 31 (average age 25), with 7 females and 5 males. On average they spent 20 minutes with the machine. The second study had nine different groups with in total 21 subjects, aged 19 to 70 (average age 38), with 12 females and 9 males. On average they spent 36 minutes with the machine.

In general, all subjects had high computer experience, their attitudes towards arts were varied, and about half had experience with children, either with their own or through working with children. All subjects were Swedish and all the quotes below are translated from Swedish.

Settings of the machine

In the first study, subjects did not have the emotion display (see below), and the machine could not be set on different influence levels.

In the second study, the machine was set up so that only certain groups saw the emotion display, some used the art postcards while others used the colored post-cards, and the condition was varied between high and low influence from the postcard on the emotion state and development of the Influencing Machine.

RESULTS

Study 1

Generally speaking, in the first study users were first curious, and then became frustrated. Often this frustration stemmed from not being able to control the machine. They had a great deal of trouble figuring out the relationship between postcard and drawings. For some users this became a barrier that stopped their interest in the machine.

Some users found the Influencing Machine drawings too simple and too slowly drawn. One user said, “[I found it] somewhat irritating as it painted so slow.” Others found this to be like “watching clouds”, thus a soothing, relaxing experience. In general, subjects did not understand that the drawings symbolized feelings, while they were more inclined to interpret the postcards in terms of emotions.

The mailbox itself was liked. One subject said: “It was so much fun, you got the urge to put down your fingers and see if there was some kind of animal in that box.” Unfortunately, the bar code reader made a beep whenever a postcard was inserted. This led subjects to think of the mailbox as a machine rather than as a form of communication with a semi-living being.

One of the goals of the Influencing Machine is to induce critical reflection in users. This was sometimes successful (“I do not know, but I shall not be able to sleep tonight I think, because I shall be thinking about WHAT the whole point was really? I do not know why really... This issue with the possibility to influence it and that. One will think about this for a long time”) and sometimes not (“I shall sleep well!”).

A complication is the frustration that users often developed with lack of control. Many users got irritated and frustrated when they could not figure it out. These thoughts and observations led to the following system design changes.

Design Changes from Evaluation

Users were confused about the emotional meaning of the imagery. Sound output was added, which helped to clarify the machine’s interpretation of input cards and its emotional state. In addition, an internal emotional display was

developed showing the level of each of the internal emotions, see figure 5. Although we were reluctant to show these internals, by offering the user an opportunity to understand how the machine is designed to feel, users can and do engage in critical reflection about whether they believe that the drawings actually express the stated internal emotion state. This display can be set to fade away over time, supporting users through their initial exploration without constraining further interaction. The system can be shown either with or without this display, allowing for further experimentation.

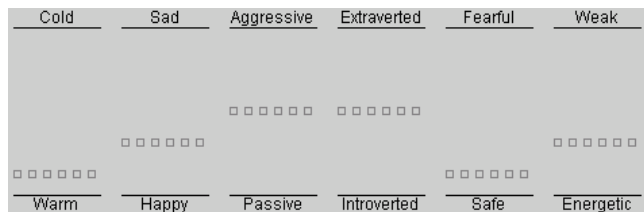


Figure 5. Part of the emotion display.

Users were also confused about the nature of influencing versus controlling the system. With improvements to emotional expression, including instant sound feedback through the music instead of mechanical Mailbox beeping for changes in emotion, it was hoped that users would have a better understanding of how they affect the system.

Finally, users were sometimes bored by the drawings themselves. This was addressed by speeding up the drawings, reducing the persistence of behaviors so that new forms appear more quickly, and adding some more complex drawings. Also, transitions between drawings needed to be handled more gracefully. In the first version, the system draws for a while and then clears the screen and starts over. After the user study, the graphics were re-implemented to remove these rough breaks by layering over one another and gradually fading away.

Study 2

The results from the second study are described by giving short-hand descriptions of the interpretative video-transcriptions for three groups. Due to space limitations, we only describe three groups here.

Description of the groups

#3 Two female computer scientists

The two subjects were 30 and 38 years old, used art postcards and had the emotional display off. It was a very active group; they put a lot of cards in the machine and tested several theories. They looked carefully at the cards that they put in the machine and then looked at the response from the machine. They sorted the cards – violent cards with pictures of an execution or hell vs. calm cards. Then they carefully looked at the response from Influencing Machine.

While they were sorting the cards the machine restarted. They tried to make the Machine feel happy, by putting in harmonious cards, and they felt that the colors of the ma-

chine’s drawings became happier, but the pictures became smaller. They tried to put in dark (night) pictures to see if the machine would respond with darker colors. The color did not get much darker, but instead they found that the response was that the Machine played saxophone music. They tried to use only violent cards to see the response from the machine, but again they felt that it was easier to understand the reaction from the machine through the music rather than through the drawn pictures. The music at this point became very dramatic.

#5 The enthusiastic quartet

The fours subjects were 20, 22, 23 and 21 years old, used colored postcards and had the emotional display on. The group almost immediately discovered the emotional display, and looked for a connection between the colors of the cards and the changes in the emotional display and the drawings. They tried to use only red colored cards to see the effects. They then tried different combinations with cards and colors. Then they tried putting every card inside the machine. They even tried inserting two cards at the same time with the UPC-codes in opposite directions to irritate the machine. This action did not affect the Machine.

After about 20 minutes the machine restarted, and then they saw the emotional display again (it had faded during the previous interaction cycle). They then actively tried to influence the machine though putting in different colors that they thought would influence the different variables in the emotion display. For example, to provoke the variable PEACE, they inserted a green card. They hypothesized that the color the machine currently uses for the drawings would also be the color that the Machine wants them to post. They then found that when they put in a red card the response from the Machine was that the emotional display changed to WARM. The Machine started to draw more quickly and with a red color and the music got more dramatic. They thought that it was a little too angry, so they tried to make it more peaceful by inserting a pink-colored card and they then found that the machine got less angry.

They deliberately tried to make the Influencing Machine express certain emotions through the drawings. Two of the group members believed that it was only a coincidence that the machine expressed the expected emotions, while the other two “knew” that their actions influenced the Machine and they tried to convince the others that they were right: “Look what happens when I put in a pink-colored card. Now it draws a sun, it has to be happy. Now we take something that makes it calm – green – that ought to make it calm. When we give it a blue card, it draws with a blue-colored pencil. If we take a card with three different colors it draws with a yellow color. Now it has started to scribble again... That is not so surprising, we put in a lot of strange colors all the time and it has not a chance to finish what it has started.”

They then sorted the card by colors and inserted them. They discussed whether the cards affected both the draw-

ings and the music. Then they completed the session by putting in all the cards. The music got very dramatic: “Now it is not in a good mood, obviously.”

#6 *Two teachers and a husband*

The three subjects were 61, 65 and 42 years old, used art postcards and had the emotional display on. The two women did not look very carefully at the cards that they put in the machine. Nor did they analyze what was happening on the screen. The machine restarted after 3 minutes. Both women kept on entering cards very quickly. The man was quiet, kept to the background, and only gave away something of his theories after about 8 minutes. In general, one woman was quite dominating and the man had a hard time convincing her that his theories could be proven. The two women realized that the machine kept on drawing even when they did not put any cards inside the machine, and thus the man’s theories could be dismissed.

The man did not give up, but discussed the emotional display and said that one has to put a card inside the machine in order to make the values in the emotional display fluctuate. He got some positive feedback on his theory from the machine, and albeit reluctantly, he got the two women to take part in some more theory forming. Unfortunately, the machine did not react to the postcard that the dominant woman inserted, at least not visibly. The man got reactions to his postcards, which in turn made him think that the machine only reacted on him. He suspiciously turned around, staring at the video camera, wondering whether this was in fact where the “control” was placed.

During this, the dominant woman made an interesting comment: she pointed at the computer under the table with the table cloth, and asked the man whether this computer was in fact connected to the machine. She meant that if it were connected, then the Influencing Machine was just a computer – not a machine in its own right. It seemed that to her a computer cannot be what she perceives that the Influencing Machine is (according to the man’s theories). If it is a computer, it must be predictable, not influenced by them.

They stopped putting in cards for a while which caused the drawings to change color until they were white and the machine restarted. They put a few cards inside the machine and then they waited for it to restart again, just to see if the drawing would change color to white again before the machine restarted. Again, the man argued that the cards they put in the machine seemed to be influencing it, but the other two argued that the card is not important and that the machine just went around in a cycle: “placed on ‘repeat’”.

They waited for the machine to restart a third time, to check if the machine would start drawing even if they did not put in a card, and they found that it did. They discussed whether the machine would restart if they stopped inserting cards or if it restarts after a certain time interval. They speculated about whether the emotions were connected with certain colors in the drawings. Finally, the dominant woman concluded that it was entirely random, while the

man kept on insisting that there were certain relationships to his actions.

In general, they did not stimulate the machine very much and therefore the drawings did not become very advanced. They only saw the circles and strokes.

Summary of session behaviors

In summary, seven of the nine groups had different theories that they tested during their session with the machine. They tried to make the machine respond in a particular way by putting a certain card or a specific category of cards inside the machine; for example, they tried to use only dark-colored cards in order to see the response from the machine. The groups that tested several different theories during the session seemed to have more fun during the session than the other groups, but after a while most of them got frustrated when the response from the machine was not what they expected. The two groups that did not test any theories, were two groups with elderly users.

Interviews

Subjects were interviewed in groups and thus sometimes discussions arose rather than mere replies to questions.

Did the Influencing Machine raise any emotions with you? Which/what kind of emotions?

A mixture of feelings is expressed by most groups: joy, frustration, irritation, curiosity, depression or happiness. Several of the groups discussed situations in which their own emotions followed the (imagined) emotions of the Machine: “Frustration, joy. Depressed by the music, one wanted to influence it, but that was not possible. When there was birds chirpings it was good.” (#5)

Group #9 had a very gloomy experience as they managed to get the Influencing Machine into a very depressed state: “You become influenced by the music, then a little bit by the colors when it all became very dark.” (#9)

Did you experience that you could influence the Influencing Machine in any way?

Two groups felt that it was not possible to influence the machine (#4 and #8). Both these groups had the emotion display off, and group #8 had low influence.

The other groups all felt that they influenced the machine, but none claimed to have figured it out entirely. One group said that they only influenced the emotion display (#1), others that they only influenced the music (#3, #8), others that they influenced the drawings (#5, #7, #9), and finally, one group believed that the music, drawings and emotional display were displaying the same thing (#6): “When we did not insert any cards, all the colors faded, the staples disappeared and the music disappeared... It was some kind of cycle, but we could influence it to some extent through continuing to insert cards.”

Where you able to read any developmental stages from the drawings?

Groups #1 and #4 did in fact mention a child in their description of the development: “I was waiting for a head-

foot, but none came. Thinking of children. They did get more advanced, in retrospect. The smaller signs were not there at the beginning. Then there were only yarn balls and lines. Yes, by and by, there were stars.” (#1)

“Yes, that I can say that I did, because it started with only lines, and then it made a circle, but longer than that did this child not come... if it wasn't for that it was not supposed to develop further, that it was beautiful as it was.” (#4)

What did you think about the sound?

The subjects described the sound through expressions such as: Skönberb and Bo Nilsson, ripples, technical museum, film music, ballet, square, artificial, improvisation electronics, improvisation jazz, Bergman's "Persona", and "from peaceful, slow to fairly heavy and loaded".

What is the Influencing Machine according to your views?

Subjects gave varying answers to this question. The best match with the original intentions was by group #5: "I think that it is a small child that draws what it feels, and if you insert colors into it, its views and thinking gets influence (it sounds really strange). It said so at the top: Internal emotions. Yes, exactly. If it got pink, it got happy, but at the same time somewhat cold and then you had to insert yellow to make it warm, but then it became less happy. I really want to see that there is a point to what you do. If you test something you want it to be usable. Maybe along the lines of helping someone to express emotions, but I did not see any logical "build up". It was some influencing machine that if you have a blockage, you can turn on the IM and get inspired." (#5 The enthusiastic quartet)

An answer from a group that felt more frustrated was: "A research tool. I experienced it as a machine to test your intelligence designed for monkeys. Only something for researchers, absolutely useless." (#1)

Group #3 provided an answer somewhere between these two extremes: "I got associations as to when you were in arts class as a kid and was supposed to draw to music. I wanted it to portray something. That you can put sound and pictures on your emotions... Then you could get something concrete. It looks like that and sounds like that. Don't know whether it has any utility." (#3 Two female computer scientists)

Several of the other groups (#2, #4, #6, #8, #9) referred to museums or art installations.

Metaphors used

As can be seen from the quotes above, recurring metaphors used by our subjects to describe their experience were museum or art installations, movies, and similarities with children. The music was described as technical museum style or improvisation in jazz.

Impressions

The results from this second study showed that the design changes did indeed achieve the desired result. Users were more positive, less confused, and almost all of them started generating theories of what was going on in the machine

and became willing to discuss the intended provocation. The subjects from the second study also used the Influencing Machine twice as long on average than the subjects from the first study. But there were still subjects who experienced frustration and who were less inclined to "get the point." They would see the interaction with the machine as random, not influenced by their behavior.

There were too few subjects to draw any definite conclusions on the settings of the machine, but the replies to the interview questions and the interactions the groups had with the machine indicated that the group who had the emotional display on did more easily grasp that the machine expressed emotions and could be influenced. We cannot claim that we could find any consistent differences between those with low and high influence of the machine. Color postcards seemed to work better than art postcards. This was a surprising result for the artists, since the color postcards were harder for users to understand than the art postcards. It seemed like the color postcards forced users to consider the interaction more carefully.

An issue that seems to be prominent in the design of affective interaction is *timing*. In the Influencing Machine, the *timing* of emotion change and development, drawings, and system's reactions to inserted postcards is key. The interaction cycle must be slow enough for users to recognize the emotions, but fast enough to attract and keep the users' interest. The intent is not for the user to control the machine, but also not to make users too frustrated when they cannot control it at all. The second study showed that the design of the machine was closer to a reasonable balance point.

The three groups that consisted of elderly people seemed to be more frustrated by the fact that they could not control the machine and fully understand how the machine worked. We had the impression that the elderly more than the younger participants felt a little stupid when they could not control the technology. The groups with younger people often tested out different theories and seemed to enjoy the session at first, but after a while when their theories did not let them to get in total control of the machine, they too became frustrated. The reason for the frustration seemed to be that they thought that there was a problem that they had to solve. When they did not fully understand how the Influencing Machine worked and when they could not get in total control of the Machine they thought that they failed an intelligence test.

In general, users' background and personality seemed to make a substantial difference in how they interacted with the machine. Users who were more inclined to form theories inserted postcards more carefully into the machine, which in turn made the Influencing Machine behave in a more interesting way. Users who were less inclined to form theories of what was going on would either insert lots of postcards in random order or not insert many at all. This in

turn made the machine either behave in ways which seemed random or get into a depressed, gloomy state.

The evaluation surprised the artists in that it made clear how intellectual emotional interaction with the Influencing Machine is. Nelly Oudshoorn has noted that system-builders tend to use an “I-methodology”, i.e. when they design for users they are subconsciously designing for themselves [12]. After the evaluation, the artists realized that, for them, intellectual and emotional experience go hand in hand, and that they had unwittingly designed the system to reflect this. If they had realized this ahead of time, they would have designed the system differently, e.g. with a tangible interface using textures to communicate emotion.

HCI EVALUATING ART, ART EVALUATING HCI

While HCI techniques can be applied to art, we do not want to force the arts into following HCI principles. Instead, we want to develop an understanding of some of the ways in which HCI and art can productively come together. Here, we are interested in how to adapt usability techniques, goals, and methods in order to be more compatible with the goals of artists. We have focused on improving the communication through the artwork, at the cost of other artists’ issues. This will not be appropriate for all artworks.

In interactive arts, the tendency has been to avoid user studies altogether, but we found they can help artists. Laboratory evaluations helped us uncover problems in interaction design: “is this interaction cycle right? how is the timing? do users understand the affective expressions?” In the case of the Influencing Machine this meant reaching the balance point between control and complete randomness (in the eyes of the users), finding good timing so that users are captivated (and not bored), finding the right level of interesting drawings, and getting better sound.

We focused on ways to help artists that want to express themselves through an interactive system to make the interaction work as intended – to help artists to get the interaction to a state where the message can be seen at all. In order to do so in a way compatible with artists’ interest, we began to move away from a decontextualised interpretation of what is going on in the studies. This study is an explicit attempt to *not* avoid the messiness of having several users together in the lab, interpreting their behaviors based on some subjective understanding of their personality and attitudes. Through such a study, we can give artists a *grounded feeling* for what works.

Here we have shown how adapted evaluation techniques can be useful for interactive art by fine-tuning interaction design. The question “is it good interaction?” can be answered, but not “is it good art?” If we want to know “is it good *interactive* art?,” we may need to more fully integrate the perspectives of art and HCI. We suggest this may be done by a ‘system critic,’ who analogous to a literary, movie, or art critic is specialized in understanding the social, cultural, and intellectual context of the system and

who simultaneously can evaluate the system using variations on standard HCI techniques.

ACKNOWLEDGMENTS

This work is part of the EU Safira Project. The Influencing Machine was built at Fraunhofer IMK.MARS. Thanks to Monika Fleischmann, Wolfgang Strauss, Boris Müller, Gabriele Blome, Thorsten Joachims, and Weidong Geng for their work on the system, to the anonymous reviewers for their helpful comments, and to Jane Austen for the title.

REFERENCES

1. Beyer, H., & Holzblatt, K. (1998). *Contextual Design: Defining Customer-Centered Systems*, Morgan Kaufmann Publ.
2. Dumas, JS, & Redish, J. (1993) A Practical Guide to Usability Testing, Ablex, Norwood, NJ.
3. Dunne, A. & Raby, F. (2001). *Design Noir: The Secret Life of Electronic Objects*. Basel, Switzerland: August / Birkhaeuser.
4. Garabet, A., Mann, S., & Fung, J. (2002) Exploring Design through Wearable Computing Art(ifacts), *CHI 2002, Interactive Poster: Fun*, 634-635.
5. Gaver, W. & Dunne, A. (1999). Projected Realities: Conceptual Design for Cultural Effect. *CHI 1999*.
6. Gaver, W., Hooker, B., & Dunne, A. (2001). *The Presence Project*. London: Royal College of Art.
7. Höök, K. (1998). Evaluating the Utility and Usability of an Adaptive Hypermedia System, in *Journal of Knowledge-Based Systems*, 10 (5).
8. Höök, K., Persson, P., & Sjölander, M. (2000). Evaluating Users’ Experience of a Character-Enhanced Information Space, *J. of AI Communications*, 13 (3) 195-212.
9. Lakoff, G., & Johnson, M. (1980) *Metaphors We Live By*, The University of Chicago Press.
10. Mateas, M. (2001). Expressive AI: A hybrid art and science practice, in *Leonardo*, 34 (2), 147-153.
11. Nardi, B. (1996). *Context and Consciousness: Activity Theory and Human-Computer Interaction*, MIT Press.
12. Oudshoorn, N. (1996) Gender Scripts in Technologie: Noodlot of Uitdaging? *Tijdschrift voor Vrouwenstudies*, 17(1): 350-367.
13. Penny, S., Smith, J., Sengers, P., Bernhardt, A., & Schulte, J. (2001). Traces: Embodied Immersive Interaction with Semi-Autonomous Avatars, in *Convergence*, 7:2, 47-65.
14. Sengers, P., Liesendahl, R., Magar, W., Seibert, C., Müller, B., Joachims, T., Geng, W., Mårtensson, P., & Höök, K. (2002) Enigmatics of Affect, Proc of *Designing Interactive Systems*, ACM Press.
15. Weiser, M. & Seely Brown, J. (1996). Designing Calm Technology, in *PowerGrid Journal*, 1(1).