

The sensual evaluation instrument: Developing a trans-cultural self-report measure of affect

Katherine Isbister^{a,*}, Kia Höök^b, Jarmo Laaksolahti^b, Michael Sharp^a

^a*Department of Language, Literature and Communication, Rensselaer Polytechnic Institute, 110 8th Street, Sage 4208, Troy, NY 12180, USA*

^b*Swedish Institute of Computer Science, P.O. Box 1263, 16429 Kista, Sweden*

Available online 9 January 2007

Abstract

In this paper we describe the development and testing of a tool for self-assessment of affect while interacting with computer systems, meant to be used in many cultures. We discuss our research approach within the context of existing cultural, affective and HCI theory, and describe testing of its effectiveness in the US and Sweden.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Affect; Culture; Evaluation; Self-report; Nonverbal

1. Introduction

In recent years, there has been increasing interest among HCI practitioners in the role of affect in the interface, both in terms of designing systems (Picard, 1997; Norman, 2004) and in terms of assessing their effectiveness with users (Isbister and Höök, 2005). Evaluation of user affect is a domain that is not as well articulated and explored as is assessing whether a system is usable, or whether it actively increases work productivity. Emotions are notoriously difficult to measure (Bentley et al., 2005). Giving a post-interaction questionnaire does not capture the richness of the emotional ebb and flow during interaction. Analyzing videotape of users for facial expression and other cues of emotion is time-consuming and difficult. Typical ratios of analysis to recorded time can range from 5:1 to 100:1 or even worse (Mandryk et al., 2006). Using biosensors to monitor physiological signals such as heart rate and skin conductance has become more popular (Picard and Daily, 2005), but offers the difficulty of interpretation of signals—for example, it is not possible to tell the difference between joy and anger (Bentley et al., 2005, p. 4). Think aloud methods may run into problems of observer expectancy

and social conformity effects, or a reluctance to share emotional information, as emotion is a highly socially reactive phenomenon (Bentley et al., 2005).

As researchers, we saw a need for further explorations of this problem space—in particular, we were interested in how to obtain information about a user's affect that could have a helpful impact on the design process itself. We are working within a tradition that encourages early and rich engagement with users during the design process, to support better ideation and evolution of the final designed interface (e.g. Muller et al., 1993). We asked ourselves what sorts of methods might enable designers to get a rich feel for user emotions as they unfold during an experience? What sorts of methods might be lightweight enough to use early and often in the design process? We worked with a bias toward self-report on the part of the user, rather than physiological data to be analyzed later on, for two reasons: (1). We wanted the dialog between designers and users to be explicit and mutual. (2). We wanted to allow for ambiguous and evocative communication between user and designer, to preserve the rich and multi-layered feelings that users might be having and to create sparks in the designer that could lead to powerful design insights—what Gaver and Dunne might refer to as ‘inspiration cues’ (Gaver and Dunne, 1999, p. 602).

We also hoped with our work, as a bi-cultural research team embedded within a European Union Network of

*Corresponding author. Tel.: +1 518 322 7569; fax: +1 518 276 4092.

E-mail addresses: isbisk@rpi.edu (K. Isbister), kia@sics.se (K. Höök), jarmo@sics.se (J. Laaksolahti), sharpd@rpi.edu (M. Sharp).

Excellence devoted to the study of affect in computer systems that includes 33 partners which represent 14 countries, to develop a tool that could be used readily by designers from various cultures, and/or by designers traveling to work with users in cultures other than their own. This has led us to an examination of the literature on culture and affect, to determine whether it even makes sense to pursue a trans-cultural measure of this nature.

1.1. *Considerations of culture and affect*

Reviews of the research on affect and how it manifests among different cultures presents a complex picture of whether there are ‘universals’ among cultures when it comes to affect (e.g. [Berry et al., 1992](#); [Kitayama and Markus, 1994](#)). One set of thinkers, with a focus on biological properties of human beings, begin with the notion that as humans share fundamental biological processes and responses to the environment, emotions can be framed as fundamentally physical responses to stimuli that emerge from these biological underpinnings. Thus they see affect as a universally human phenomenon that is more similar than it is different, across cultures. Among the most well-known proponents of this view are Ekman and his research collaborators, who have conducted extensive examination of identification of facial expression in many cultures ([Ekman et al., 1987](#)). These studies have led to a consistent cluster of identifiable emotions said to be universal (though this claim is not without its important detractors—for example, [Russell, 1991, 1994](#)).

In contrast to this position are those who conceptualize affect as entirely socially constructed. This group of researchers, among which the preponderance are cultural anthropologists and linguists, argue that a person’s affect is profoundly determined and situated by his/her enculturation and context, and that we can see evidence of this in language and in the study of particular cultures and their idiosyncratic emotion concepts. For example, [Wierzbicka \(1992\)](#) describes the notion of ‘song’ among speakers of Ifaluk (a language spoken on an atoll in the Caroline Islands) as a unique emotion without equivalents in other cultures.

In-between these two poles are those who feel there may be some universal qualities of affect that are shared, with a strong dose of cultural and individual factors influencing how affect is experienced and expressed. One example of research to support this middle-ground is a large-scale questionnaire-based study across 37 countries conducted by [Scherer and Wallbott \(1994\)](#), which showed large degrees of consistency among these countries in the experience of 7 major emotions, as well as significant degrees of cultural variance in ‘emotion elicitation, regulation, symbolic representation, and social sharing’ ([Scherer and Wallbott, 1994](#), p. 310). (For an interesting review of the divergence of beliefs and methods between the psychological and anthropological fields and an excellent

attempt at setting integrative goals, see [Norenzayan and Heine, 2005](#).)

We were interested in determining which, if any, proposals for universals seemed particularly stable or promising. We found two researchers who seemed to converge upon a couple of universal looking qualities of affect that were of use to us in our thinking. One was [Wierzbicka \(1999\)](#), a linguist who analyzed emotion words in many languages, looking for common patterns. She found that some concepts were included in every language, one of which was the general notion of positive/negative. This result was corroborated by an analysis by Osgood and colleagues ([Osgood, 1977](#)), who conducted an extensive examination of ‘subjective culture’—how people within a culture view themselves and evaluate their way of life—in 30 communities spread around the globe. They found that all of these communities used dimensions that could be categorized (in English) as valence (good/bad), potency (strong/weak) and activity (active/passive). Aspects of this common-dimensional space have been corroborated by other emotion researchers, such as Russell, who created a dimensional space with valence (good/bad) and arousal (high/low energy) as the axes ([Russell, 1980](#)). (It’s worthwhile to note here that Osgood constructed an ‘atlas’ of these universals, along with what he called ‘sub-universals’ and ‘uniquenesses’—his research uncovered qualities at both ends of the universalist and social constructionist poles.)

Several researchers have emphasized the fundamentally social and dynamic nature of affect (e.g. [Averill, 1980](#); [Wierzbicka, 1999](#)). If one takes a component-based view of emotions (e.g. [Mesquita et al., 1997](#)), one can see that goals and other factors have a heavy influence on what is expressed. We cannot truly separate expression of emotion from its context, social and otherwise.

After considering these analyses of culture and affect, we would state the following as guideposts for our work:

- There seems to be some physical grounding of affect that is universal (evidence in facial expression research, in linguistic analysis, in Osgood’s work, as well as in Scherer and Wallbott’s study). Osgood’s universals are a helpful place to begin.
- This is heavily overlayed with and influenced by cultural/social/individual context, which is crucial to understanding and engaging in an emotional dialog with users.
- Emotion is a social and dynamic communication mechanism, not just a ‘knee jerk’ physical reaction. Therefore we must consider the form and particulars of the ‘conversation’ when constructing our instrument.

This review has given us hope that we can create a self-report instrument that makes use of some basic common dimensions, while allowing for a range of situated expression.

(It is of course important to note that the design of such an instrument to aid design would not fully account for nor address differences in how various cultures experience and make use of interactive technologies and media, which is an entire research area of its own).

1.2. *Why a nonverbal, sensual measure*

Our research team wanted to avoid eliciting words, in order to access aspects of a person's affective experience that might not be readily processed into language as s/he interacted with a system. Recent research in decision-making has shown that people make intuitive decisions using affective processing, before they are able to verbalize what and why (Myers, 2002). Also, parts of affective processing are taking place in areas of the brain not readily accessible to language (LeDoux, 1996). Language is wonderful for summarizing and categorizing and processing one's emotions after an experience, but might sometimes be a clumsy tool for communicating affect in the fleeting moment of interaction, particularly if one is feeling a jumble of half-resolved emotions-in-progress. We hoped using the visual sense as well as the sense of touch would elicit different kinds of responses from users, and perhaps allow for greater unresolved ambiguity in their expression. (Thus providing designers with richer fodder for iteration.)

We also thought it was possible that using objects would somewhat relieve the social conformity impulse in the user. If s/he did not have to 'say' that the interface was annoying

but could instead show it, then such feelings might be allowed to surface more frequently and be communicated.

Also, we are committed as a research team to making affective evaluation engaging and interesting for users. We felt that 'playing' with objects would be more fun for users than filling out a survey or thinking aloud, and might thus lead to more relaxed and creative responses.

It's important to note here that there has been some precedent in attempts to develop nonverbal, cross-cultural measurement instruments to aid in product design. The most closely related example that we could locate was a study of an animated character used to conduct product evaluations, called PrEmo, short for Product Emotion Measurement instrument (Desmet, 2004). This system used a pre-constructed set of 14 emotional character animations—desire, pleasant surprise, inspiration, amusement, admiration, satisfaction, fascination, indignation, contempt, disgust, unpleasant surprise, dissatisfaction, disappointment, and boredom (see Fig. 1), which were tuned especially for emotional response to product concepts such as car models. Although the authors claim some success for this instrument across several markets, the pre-set range of emotion does not allow for the rich dialog between the users and the designers of an evolving system, nor the flexibility of emotional taxonomies and categorization that were our aims.

We should note here that there is a wide range of techniques for eliciting nonverbal emotional feedback in the general social science literature (such as play therapy, see O'Connor, 2000) that is beyond the scope of this work. Our focus is specifically on enabling emotional feedback

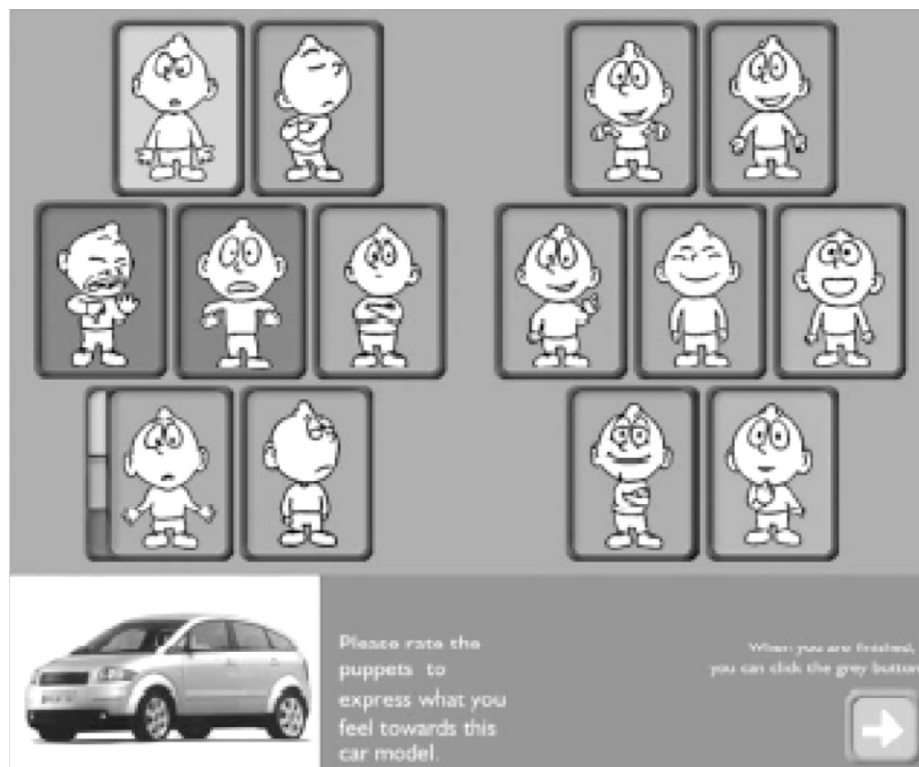


Fig. 1. Product emotion measurement instrument interface (from Desmet, 2004).

within a design context. Additionally, there are other choices that have been made in the design of nonverbal, real-time measures (e.g. anthropomorphic scales) which we considered during our development process and do not cover here—please see our CHI 2006 paper (Isbister et al., 2006) for more information about related nonverbal measures we examined.

2. Creating the SEI—an iterative process

The process by which we developed the Sensual Evaluation Instrument is described in greater detail in our CHI 2006 paper (Isbister et al., 2006). Briefly, we began with very open-ended explorations of the affective properties of objects, initially among the research team and later with a few outside participants. We initially attempted to use color as the differentiating factor among the objects (see Fig. 2), but quickly saw that individual meanings and emotional taxonomies of color were far too idiosyncratic, as well as being heavily driven by memory and narrative.

We moved from this first phase to a decision to craft biomorphic objects whose contours expressed affective qualities, not unlike those in the work of talented animators (e.g. Disney's famous flour sack example, Fig. 3). We felt that using such forms could lead to an instrument that would be intuitive and familiar at a visceral level, while still ambiguous enough to be open for interpretation.

We worked with a sculptor who had experience crafting sensual and biomorphic shapes—Rainey Straus (www.raineystraus.com). She started from a set of descriptions of emotions that were based upon our observation of users in the early prototyping phase. Here are the descriptions that she received:

- Confusion (I don't get what's going on here)
- Frustration (what the system just did drove me nuts; or, I can't solve this level and I hate this right now)

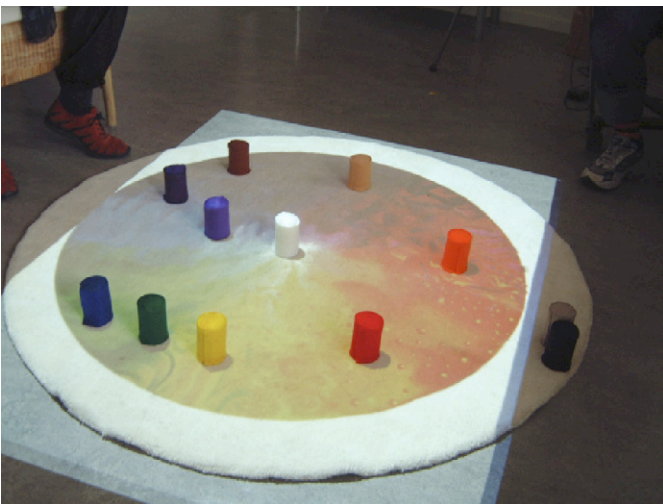


Fig. 2. Early sensual evaluation objects arrayed on the Russell emotional dimension space of valence and arousal.

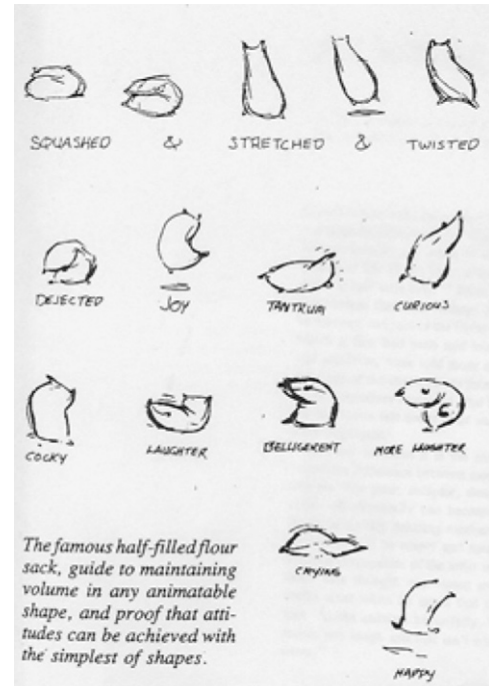


Fig. 3. The Disney flour sack, an illustration of emotions conveyed by an object despite the absence of a clearly articulated human figure (Johnston and Thomas, 1995).

- Fear (the game is making me anxious; or, I think I might've erased the wrong files)
- Happiness at success (I just cracked a level; or, I just figured out how to do a new thing)
- Surprise (positive—something good happened I wasn't expecting)
- Surprise (negative—something bad happened I wasn't expecting)
- Satisfaction (something happened that I like)
- Contentment (all is okay, going smoothly)
- Frantic stress (things are out of my control, too much going on)
- Flow (I'm in my groove right now, really enjoying working with the system, we feel as one)
- Neutral (not feeling emotions right now, just working...)

She was not asked to create a one-to-one mapping, but rather to use these as a starting point for her work. In the end she crafted eight objects (see Fig. 4). They were sculpted in a bakeable clay, then cast in a durable white plastic material that was nonbreakable and neither very heavy nor overly light. The objects are not cold to the touch—rather they warm quickly in the hand. They are smooth surfaced, but not highly glossy.

Internally, we assigned names to the objects to make it easier to code data, as follows (follow the first image in Fig. 4): back row—spiky, pseudopod; next row—anteater, bubbly; next row—stone, doubleball, ball; front—barbapapa (named after a figure from a French animated

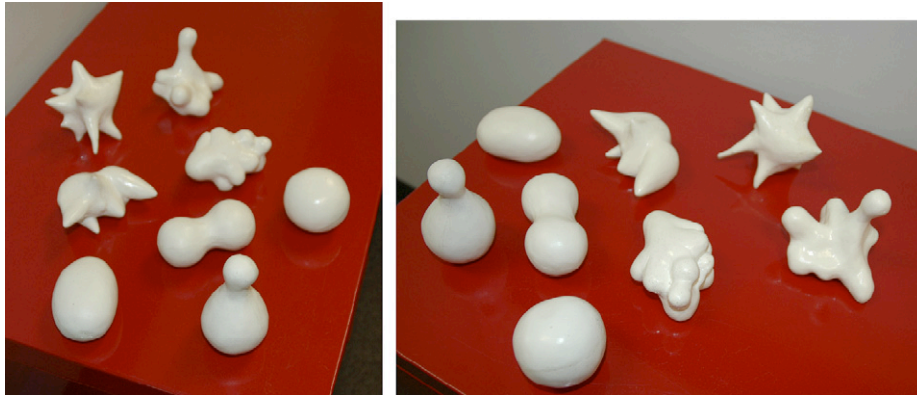


Fig. 4. Two views of the sensual evaluation instrument objects.

cartoon popular in Sweden). We did not use these or any other names for the objects with participants—these names are introduced only to aid you, the reader, in following discussion of object use in this paper.

We researched tactile and visual qualities of materials and objects, and found that there is some evidence of consistent mappings by people to certain qualities. For example, sharp objects are unpleasant and potentially dangerous, and to be avoided (Picard, 1997). Symmetry is generally pleasing (Bunch, 1989). Examining our object set, one can see that they can be clustered into sharp and more rounded groups, as well as into symmetrical and asymmetrical groups. These properties seemed to have an interesting impact on the responses the objects received (described in the Findings Section 3.3 below).

3. Testing the SEI in two cultures

After constructing the prototype set of objects, we used molds to cast multiple copies so that we could conduct studies in various locations. We designed a test that would be first conducted in the US then replicated in Sweden. We used a laptop and easily accessible software, so that we could readily follow the same process in both locations. The test was conceived as a co-exploration with those who participated, of these prototype objects. We wanted both to observe participants trying out this new method of giving feedback, and to learn from their own reflections about the process. In terms of testing across cultures, we hoped to learn from this pilot test whether the objects could ‘work’ in two different cultures, and what the differences might be between user experiences.

3.1. Method

Participants were recruited in both countries from a pool of students and researchers who were interested in HCI issues. There were 12 participants in the US (10 male, 2 female) and 12 participants in Sweden (8 male, 4 female).

We took participants through an hour-long process that consisted of 5 stages:

- *Explanation of research purpose and orientation to the objects:* In this phase, the participant was shown the objects and encouraged to handle them, and the experimenter described the purpose of the research and emphasized that any type of usage of the objects to convey affective state was fine (multiple objects at once, movement of objects, and so forth).
- *Use of objects with a subset of the IAPS (International Affective Picture Set):* In this phase, the participant was shown a series of images taken from the IAPS (Lang et al., 2005), and asked to use the SEI objects to indicate affective response to each picture. (See the Findings Section 3.2 for examples of the images used). The images were selected because their arousal/valence scores mapped roughly to the feelings that the artist initially intended to convey with the objects.
- *Use of objects with a computer game:* In this phase, the participant played through the first puzzle in a PC adventure game, *The Curse of Monkey Island*, and was instructed to use the objects to indicate affect during play. The experimenter was present during this phase, and was available to offer hints/tips on using the game during play. (Note: this game is known for its extremely funny dialog and the frustrating nature of its puzzles—so we chose it in part because we knew it was likely to successfully elicit strong emotions from players.)
- *Use of objects during a chat:* After the game, the participant was asked to chat using AIM instant messaging with the experimenter’s assistant, to discuss how it was to play the game. The experimenter left the room during this portion, after instructing the participant to use the objects to indicate affect while chatting.
- *A discussion of how it was to use the objects to give affective feedback:* At the end, the experimenter returned and walked the participant through a series of questions about what it was like to use the objects, including solicitation of suggestions for improvement of the SEI. Questions included: What was it like to use the objects to express

emotion? Did you find that you had a consistent set of mappings? How hard or easy was it to use them? Any other thoughts? Suggestions for changes or alternatives?

The sessions were digitally recorded (with participants' consent), so that we could later analyze usage in conjunction with what the participant was seeing on-screen. Recording was done in a picture-in-picture format in the US (see Fig. 5), and with separate camera angles that were then recombined in a similar way in Sweden, so that we could see what was happening on-screen, with the objects, and also see the person's face and hands (see Fig. 18 for an example of the Swedish post-study analysis video).

3.2. Findings

3.2.1. Calibration with images

Among US participants, although there were definitely individual differences in reactions, we found some consistent patterns of object use when considering the valence (positive or negative) of the calibration images. Fig. 6 shows that US participants tended to use the sharp objects (spiky, anteater) for images with negative valence, and more rounded objects (ball, stone) for images with positive valence. (See Fig. 7 for examples of images used from the IAPS.)

Swedish participants did consistently use the spiky object for negative valence images, but did not show the smooth object pattern for positive valence images (see Fig. 8).

Neither group showed clear patterns of object use when considering the arousal dimension for the photos (see Figs. 9 and 10).

We noticed with the calibration task, that participants from both cultures self-reported making object choices

sometimes due to similarities in the contours of the SEI object and what was in the photo—for example, spiky looked like the shark's teeth, or the anteater looked like the mountain range. Participants did not report this sort of shape matching during the other two interactive tasks in the study, leading us to conclude that the IAPS may not be the most reliable form of calibration for our instrument, because it relies on still images that can be analyzed purely visually rather than unfolding snippets of experience.

3.2.2. Individual taxonomies

There were patterns of use by both groups that suggested consistent underlying dimensional mappings, as well as a high degree of interpersonal variance. Several participants in both the US and Sweden indicated that the sharper shapes (spiky and anteater) were associated with anger and frustration. This association was clearest for Swedes to spiky. Because anteater had rounded edges on one side, both groups characterized it as an object that could shift

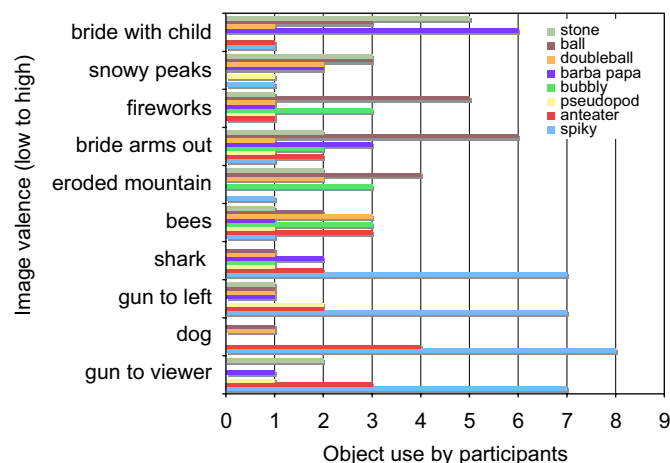


Fig. 6. Object use by US participants for each IAPS image, arranged by valence (from low to high). SEI object key is arranged from rounded to spiky, top to bottom (at right).

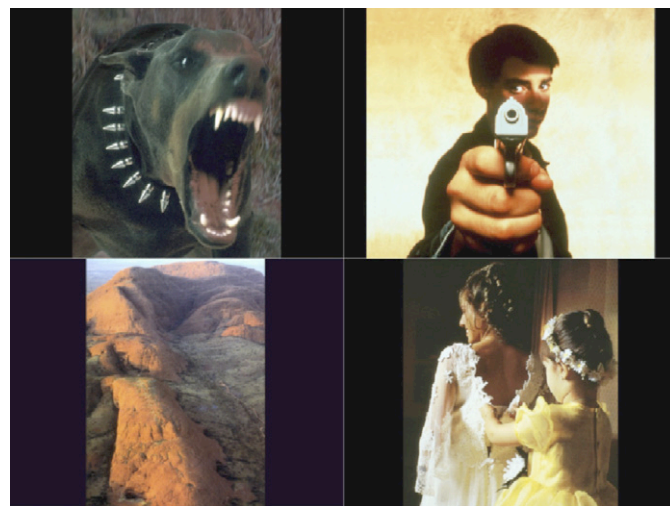


Fig. 7. Sample high arousal/low valence (above) and lower arousal/higher valence (below) images selected from the IAPS.

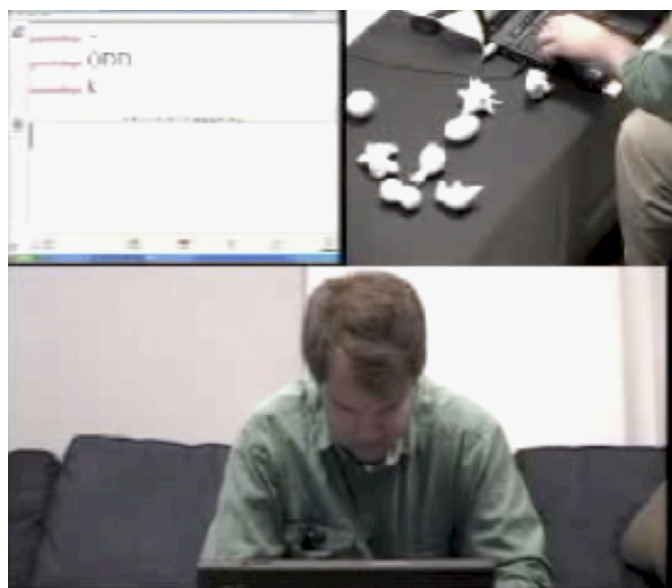


Fig. 5. In the US, we used a picture-in-picture capture process so that we could see the participant, his/her hands and the objects, and what was on-screen. In Sweden, we blended output from three cameras to achieve the same effects.

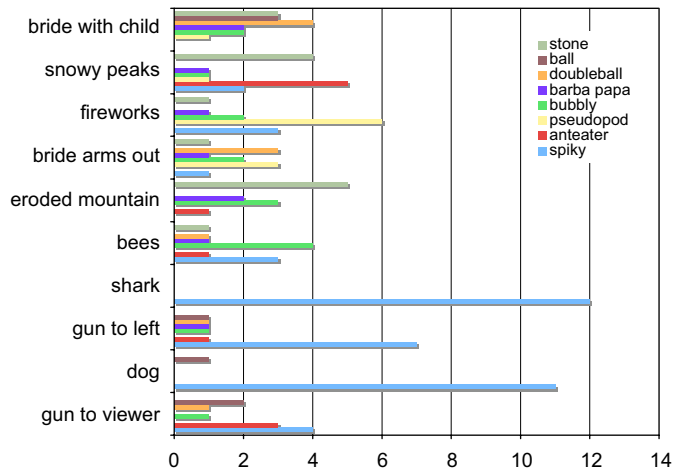


Fig. 8. Object use by Swedish participants for each IAPS image, arranged by valence (from low to high). SEI object key is arranged from rounded to spiky, top to bottom (at right).

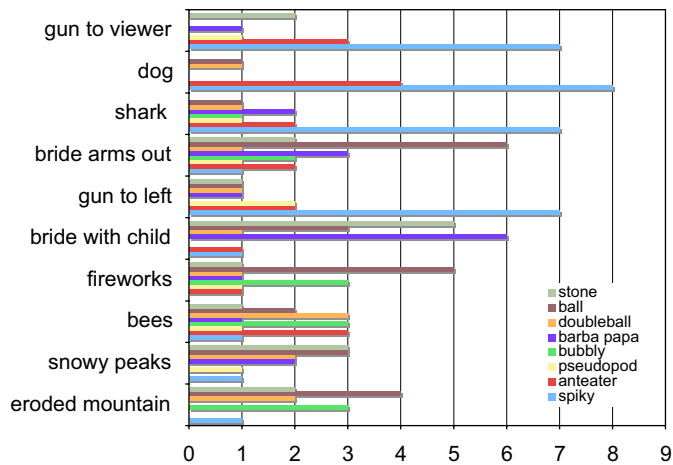


Fig. 9. Object use by US participants for each IAPS image, arranged by arousal (from low to high). SEI object key is arranged from rounded to spiky, top to bottom (at right).

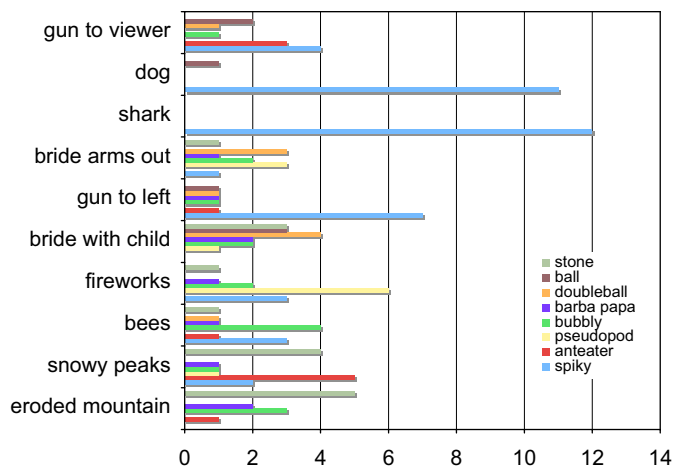


Fig. 10. Object use by Swedish participants for each IAPS image, arranged by arousal (from low to high). SEI object key is arranged from rounded to spiky, top to bottom (at right).

meaning. Some members of each group suggested malicious humor as one interpretation of the object.

Members of both cultural groups said that the more rounded objects as a category seemed to indicate more positive, calmer emotions to them. Participants in both groups noted that these objects felt very good in the hand.

Some objects had a high degree of consistency of mapping for both cultural groups—spiky (anger, frustration), ball (calm, positive), and stone (calm, positive) received consistent descriptions in the post-study discussion with participants. Other objects had consistent patterns for one group and not for the other. For example Swedes found barbabapa difficult to interpret, whereas several US participants associated it with humor.

US participants used bubbly for both humor and for confusion and frustration. Many commented that the bubbly form seemed chaotic and suggested confusion or indeterminacy. One Swede reported a similar feeling, but the rest found this shape rather confusing. Pseudopod got some similar reactions to bubbly from US participants—one person used this to indicate that he had a goal in mind but it wasn't wholly resolved—a feeling of directed anticipation. Swedes found this shape odd and confused, and one associated it with an atomic explosion cloud!

Doubleball was not used much by participants from either culture, and interpretations varied, though they usually had positive overtones. One participant in each group remarked that this shape reminded them of a peanut.

3.2.3. Use patterns

Members of both cultures did use the objects, but the Swedes used them less frequently overall than the US participants. Selection of objects during the session (total time in the game and chat tasks was typically ~30 min total) ranged from 0 to 53 times among the US participants, and from 0 to 37 times among Swedish participants. Participants used the objects to express emotion both when playing the game with the experimenter present, and when instant message chatting with the experimenter not present in the room, although usage was less frequent in the latter case—two Swedish participants, and one US participant did not use the objects at all during chat, whereas this happened only once for the game (a Swedish participant). (See Figs. 11 and 12.)

There were large differences between the participant populations in terms of which objects were used the most during the experimental tasks. Some objects, such as anteater and pseudopod, had relatively similar frequency of use, whereas other objects, such as double ball and barbabapa, dramatically differed (see Fig. 13).

The use of multiple objects at one time varied considerably among participants in the US, and was infrequent among Swedish users (see Figs. 14 and 15). Ten out of the 12 US participants used multiple objects simultaneously at least once, with total multiple object use peaking at 21 instances for one participant. Only two of the

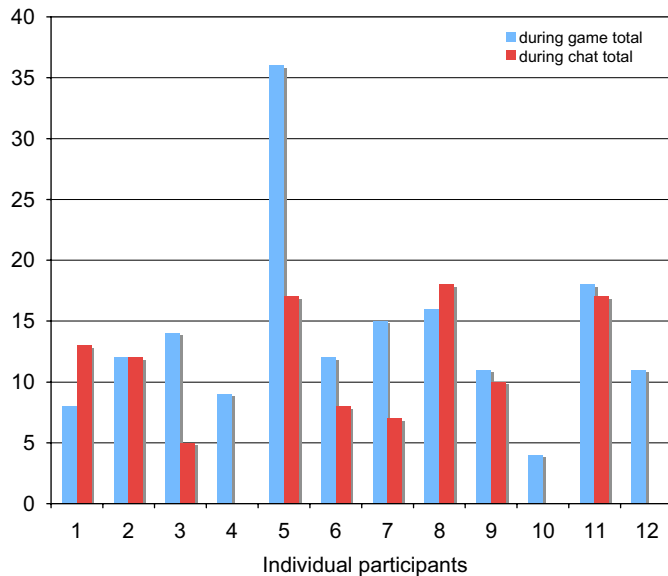


Fig. 11. US use of objects in the game and chat sessions.

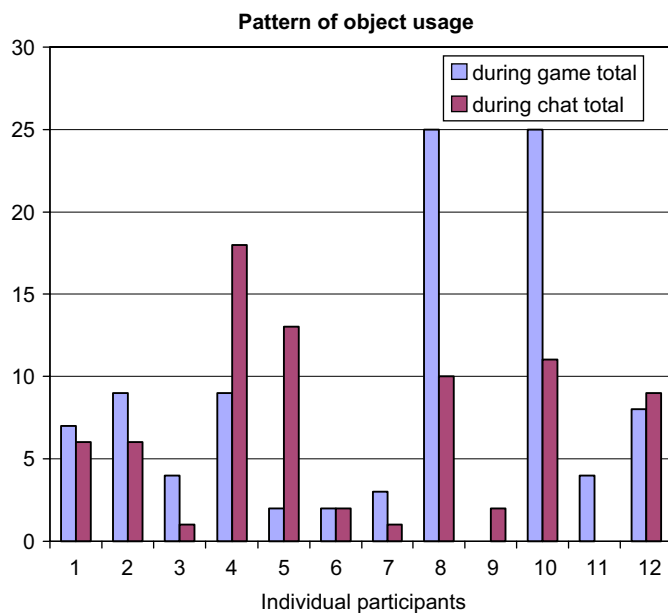


Fig. 12. Swedish use of objects in the game and chat sessions.

Swedish participants used multiple objects at once, and of these the highest instance was only 4 times.

Considering style of multiple object use, some participants held more than one object in the hand (Fig. 16) whereas others arrayed or stacked them (Figs. 17 and 18, respectively).

Participants used the objects in a variety of ways. Some held an object in one hand to indicate that was their current selection (e.g. Figs. 16 and 19); others arrayed the objects near them (e.g. Figs. 17 and 18). Some participants

(e.g. Figs. 19 and 20) gestured with the objects in addition to holding them in the hand.

Finally, there was a spread of how many of the total object set was used during the study itself. Some participants only used a handful of the objects (e.g. Fig. 17, a US participant who said he would use the more pointy objects only if he saw ‘the blue screen of death’, or Fig. 20, a Swedish participant who only used three of the objects), vs. Fig. 18, a subject who used every object except Spiky during his session. Fig. 21 contrasts two different US participants’ patterns of object usage—participant one made use of only five of the objects, whereas participant two used the full range of objects, and much more frequently. Considering the two cultural groups as a whole, there was a general trend of the Swedish participants using a narrower range of the objects than the US participants (see Fig. 22).

3.2.4. Comparison of participant comments

Participants from both cultures gave us both positive and negative feedback about the experience. Both Swedes and US participants commented that they liked holding the objects in their hands—especially the smoother ones. For example, one US participant remarked ‘They were fun to play with, just to roll them around in your hand.’ (This participant had an object in hand almost constantly during the session.) A Swedish participant said it is nice that the objects are not fragile, that you can play around with them without fear of them breaking. Several US participants said they liked being able to use more than one object at one time, and that this was more fun for them than filling out a survey would be. When asked if he would prefer to verbalize his emotions instead of using the objects during a game, one Swedish subject jokingly commented that ‘then I would be frustrated and use this one instead (spiky)’.

Both groups expressed frustration that the object set was not ideal. Both Swedes and US participants were looking for objects that could express emotions such as ‘joy’ or ‘sadness’ or ‘despair’ and did not see options among those offered to them. Interestingly, the solution proposed by most Swedes was to have fewer and more differentiated objects, while most US participants suggested having lots more objects and perhaps letting people choose from among them to create their own personal set. It’s also interesting to note, along these lines, that US participants made far more frequent use of multiple objects than Swedish participants.

Swedes and US participants both noted that the doubleball object was not very useful, and that the ball and stone seemed too similar to be usefully differentiated.

Participants from both cultures made some very interesting suggestions for evolving the instrument, which included:

- adding the ability to squeeze or shape the objects somehow with one’s own grip, to further indicate emotion,
- adding droopy objects to indicate despair or low energy level,

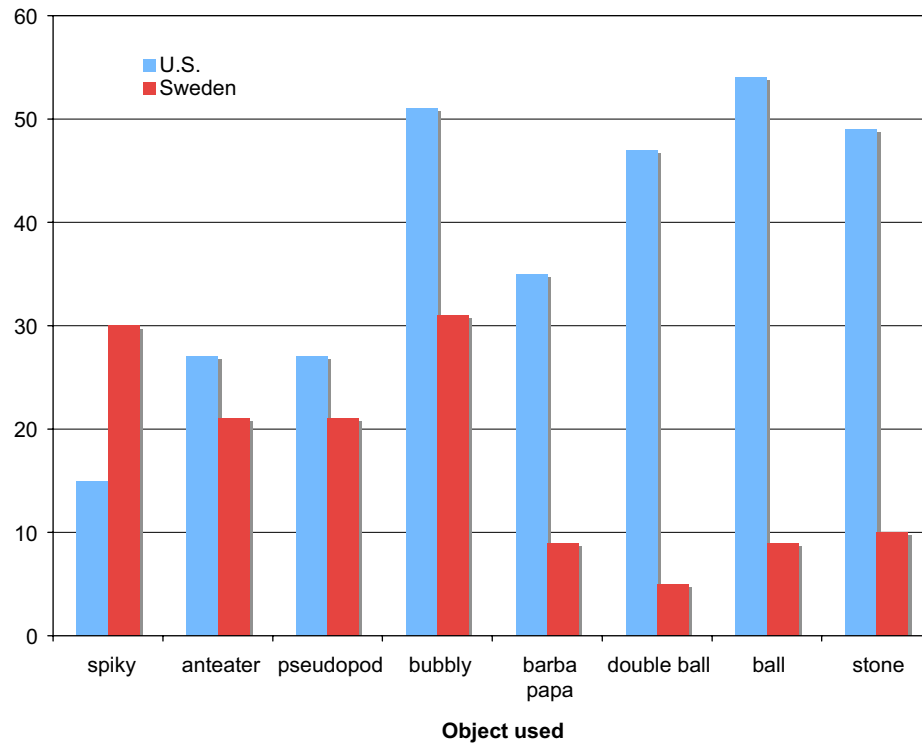


Fig. 13. Comparison of frequency of individual object use during tasks between US and Swedish participant groups.

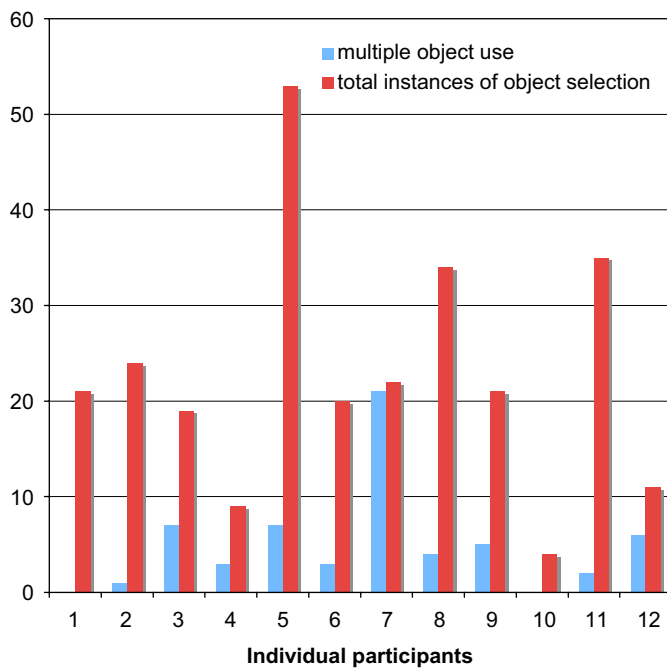


Fig. 14. Multiple object use by US participants.

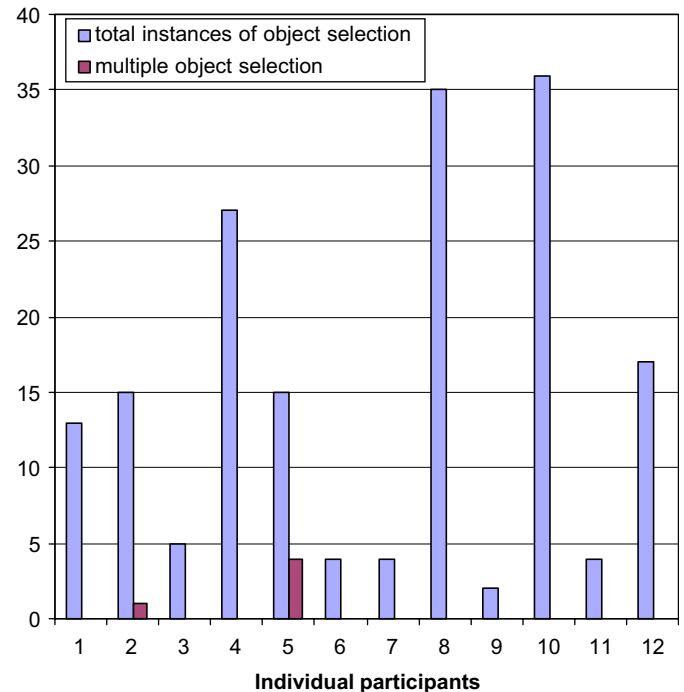


Fig. 15. Multiple object use by Swedish participants.

- introducing scale, texture, and color as variables,
- having some visible representation of the objects on screen so they are easier to remember and use without disturbing the task.

One other factor for consideration in the object use (briefly mentioned in Section 3.2.1) was the tendency for participants sometimes to match the shape of the object with the shape of the thing they were reacting to. This was

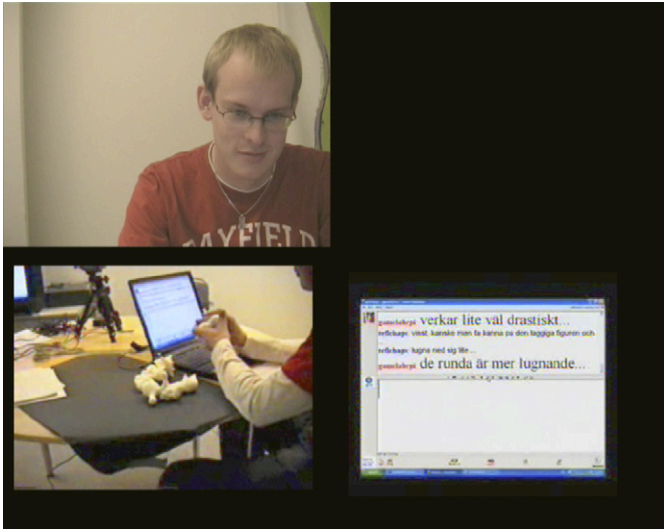


Fig. 16. A Swedish participant rolls two objects in his hands. This participant was one of the few Swedish subjects that used multiple objects. For a long time he held ball and stone in his hand, and rotated them, while chatting. He commented that for him the shape of the object is not as important as how he uses it. Hence he chose objects that felt nice to hold in the hand.

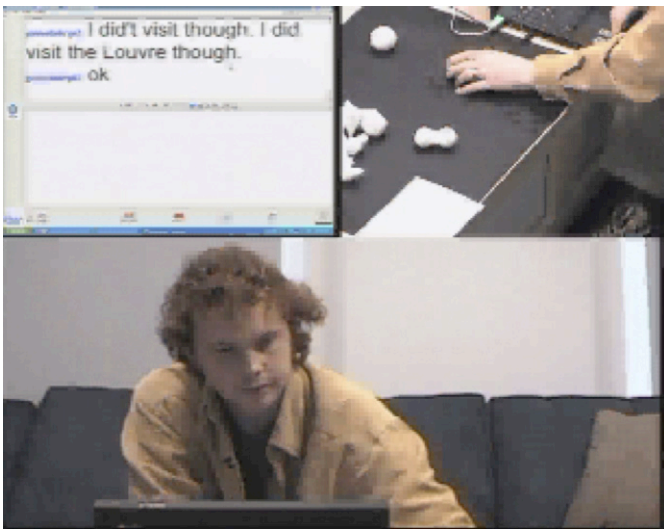


Fig. 17. A US participant arrays multiple objects beside him. This participant used multiple objects arrayed in a triangular formation to express his reaction. His use of the objects was limited to those that were most rounded in form (ball, stone, doubleball, barbapapa), and he said he probably would not use the sharper forms (such as spiky) unless he saw the ‘blue screen of death’.

especially so with the IAPS image calibration (some subjects used the pseudopod or barbapapa in response to a gun image, or the doubleball or bubbly because it looked like the rounded mountains, and so forth). This tendency seemed to become less prevalent with the active system feedback sessions, but is a concern in further evolution of the shapes themselves.



Fig. 18. A US participant has stacked barbapapa on top of stone. This participant stacked objects in various ways during his session. He used every object except spiky in his session.



Fig. 19. A US participant gesturing with spiky. He shook spiky in response to a lag problem in the game during play (and at other times). He tended to hold the objects in his hands when they were ‘active’. He made frequent and dramatic use of the objects including broad gestures—over 50 instances in the 30-min session.

3.3. General discussion of results

This bi-cultural testing process confirmed some of the initial results we found in the US study (see Isbister et al., 2006 for extended discussion):

- Using the SEI seems to be fun and engaging, in part because it makes use of the body and sense of touch, and also because it allows for flexibility in response. We observed stacking, rolling, holding in the hand, gesturing, and arranging in an array next to the laptop—a wide range of different modes of use. And we heard

positive comments from both cultural groups about the process.

- There are indications (through frequency and range of observed use patterns and comments afterward) that this is a feasible way to give in-process feedback without too much disruption, as we had hoped.

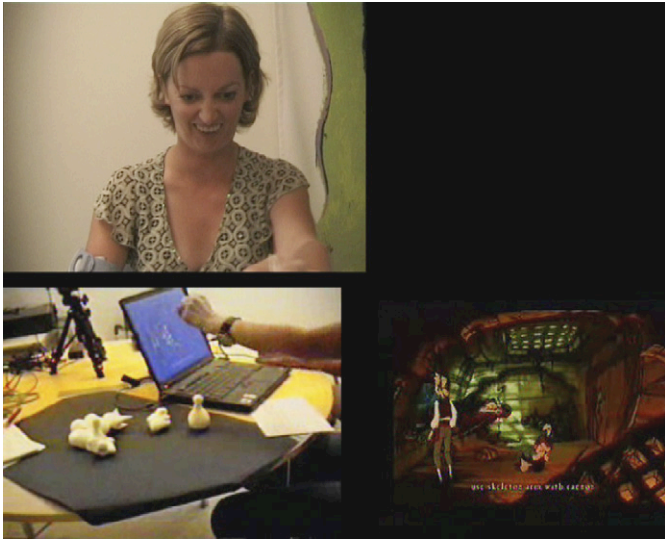


Fig. 20. A Swedish participant makes stabbing motions with spiky towards the screen. She frequently used spiky to indicate frustration when not making progress in the game or when the chat session was progressing slowly. Usually she would pick up the object and shake it before putting it down again. But sometimes this was extended with stabbing motions towards the computer screen. She only used three of the objects, including spiky.

- Use patterns and verbal descriptions also seem to indicate that this could be evolved into a reasonably consistent instrument in terms of general affective dimensions (particularly positive and negative valence), even across cultures, while maintaining a flexibility and subtlety of expression for participants (evidence for this can be seen in the wide range of usage patterns described in Section 3.2).

There were important differences in the use of the SEI between Swedes and US participants—less frequent use and range of object choice by Swedes, less meaning attributed to some of the objects than by US participants, fewer multiple-object usages by Swedes—but we feel with some tuning, our object set has a reasonable chance at supporting design feedback in multiple cultures, as we had hoped. So long as we are able to set up a rich and helpful communication between designer and potential user in a range of cultures, we will have achieved our aims with this particular instrument.

It is important to note some key limitations of this exploratory cross-cultural examination of the objects: first, the results reported are observations of trends, not statistically significant results of a large-scale comparison. Second, though the participant pool in both studies was similar in make-up (HCI researchers and students) we did not rigorously screen for demographic and personal characteristics of participants. It's possible that there were differences in the participant pools along axes other than culture that may have contributed to different use patterns. Finally, these two cultures (US and Sweden) could be

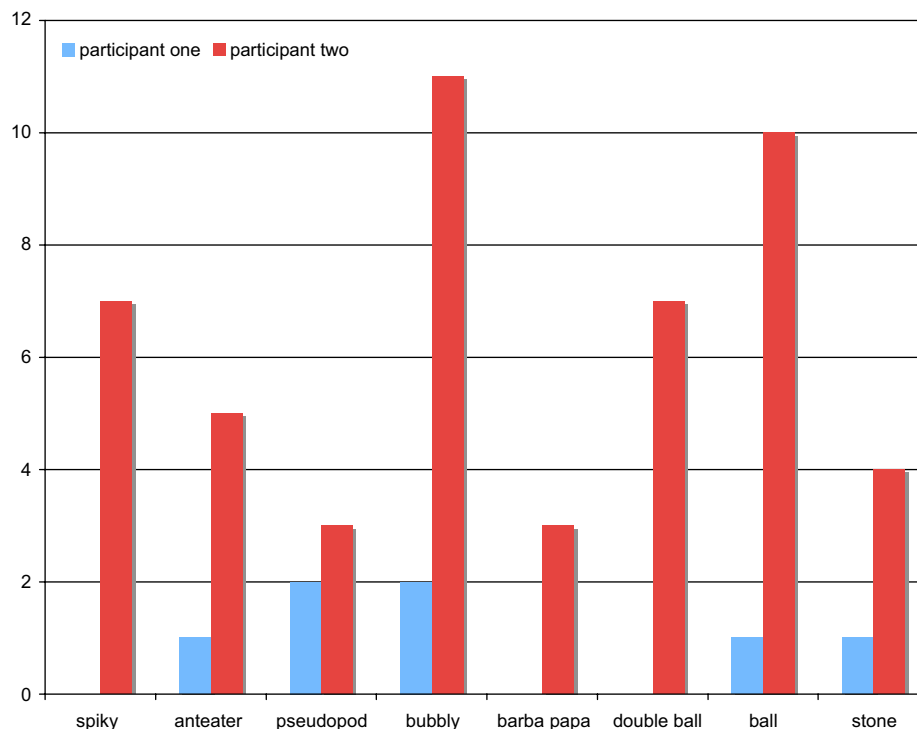


Fig. 21. Comparison of overall frequency of use of each object, between two different US participants in the study.

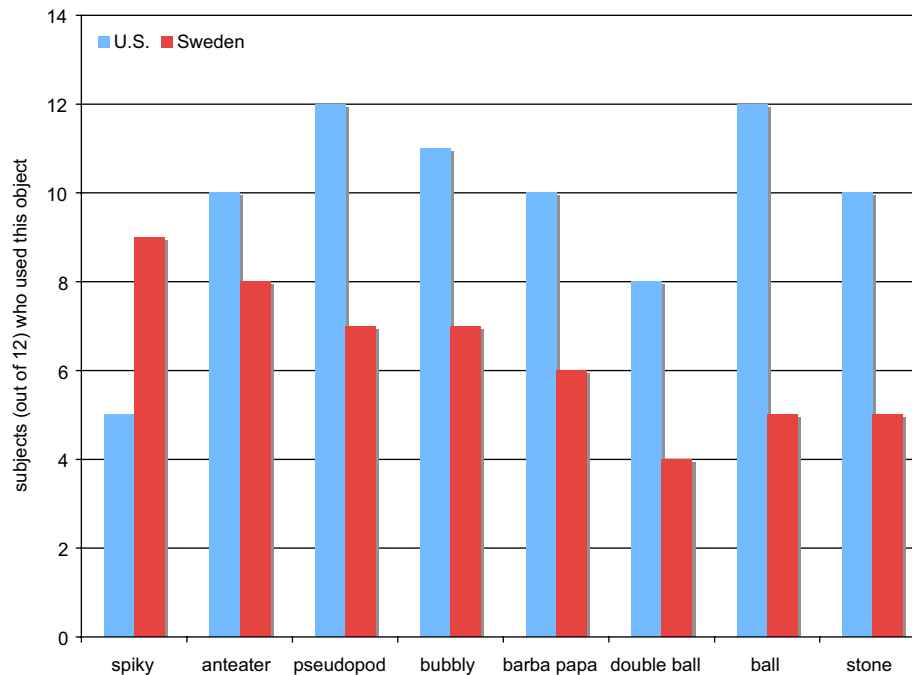


Fig. 22. Comparison of how many participants in each cultural group made use of each object.

categorized as quite similar, as opposed to, say, a comparison between a European and an Asian culture. This initial study was meant to examine the promise of the SEI as a cross-cultural instrument for garnering feedback, not as a definitive proof of its validity. We feel our results are exciting enough to warrant further exploration of the SEI as a feedback method for use across cultures, and we plan to address these limitations in future work.

In terms of use of the SEI as a feedback tool for designers, conducting these two preliminary studies provided some helpful indications about the strengths and limitations of the approach. We noticed that participants seemed to use the objects to indicate moments of emotional transition that were relatively strong, rather than as a constant, finely tuned expression of all of their subtle emotional shifts. They also tended to use the objects less frequently when highly immersed in an interaction, until a relatively strong emotion led them to ‘take a step back’ and provide feedback with the objects about what was going on. Having observed and analyzed these early tests, we feel that this method would *not* be ideal for eliciting user feedback about very time-critical and/or highly attention-demanding interfaces, as users would have neither the time nor the attention to devote to handling the objects. Nor would the SEI likely be able to provide extremely fine-grained, moment-by-moment emotional data about participants, given the use patterns observed thus far.

Instead, the instrument seems well-suited for creating a low-stress, informal and playful dialog between designers and potential users of a system that is not relentlessly time and attention demanding, which can defuse the social stigma of providing key and sometimes critical feedback

about the system verbally. The objects provide a way to avoid direct criticism and allow for playfulness in approach and more range of expression than a questionnaire or even a think-aloud might (such as several conflicting emotions, and sharing unresolved and somewhat amorphous feelings).

Examples of ideal target systems for use of the SEI (meant to engage the emotions, perhaps in new ways, and not intensely time and attention demanding) include:

- new mobile phone or other hand-held device interfaces,
- smart home controls and feedback,
- anthropomorphic or highly social (e.g. avatar-based) interfaces that use human social and emotional communication with users,
- videogames which do not require extremely time-intensive hands-on operation at all times,
- evaluation of first impressions of products and websites with a strong emotional component,
- traditional desktop computer interfaces which the designers anticipate will generate strong emotions.

The SEI did not entirely eliminate the need for words as a tool for discussing emotion. For the experimenters, the post-use discussion with participants was invaluable in helping us to best understand the rich expression that the participants showed through use of the objects—their supplemental verbal ‘decoding’ was useful as a bridge into the gestural and multi-layered way that each used the objects. However, we definitely saw evidence that the nonverbal, tactile method of providing feedback while

working with the system had merit and transcended what we could have accomplished with a think-aloud method.

In contrast to other nonverbal measures such as the PrEmo faces, or a dimensional valence (positive/negative) and arousal approach, we felt the SEI allowed for more subtle, multi-layered, and open-ended participant reactions to the interfaces that were used, and would provide a designer who observed the test (or who saw the video later) with a far richer level of access to what was happening for the user of the system. As such, the SEI seems most valuable as an alternate emotional channel between designer and user (rather than as a summative tool for others to use in guiding their evaluation of the ‘success’ or ‘failure’ of a given design).

In short, we feel the SEI is best-suited for eliciting the type of emotional feedback that may be difficult to bring into words or may have social stigma attached, as someone is experiencing a new system or product that is likely to create emotional responses, particularly those that may be multi-layered and complex. There is still work to be done in terms of establishing the robustness and efficiency of this method of eliciting feedback, but there is certainly a need for designers to receive this kind of nuanced emotional feedback as they design ever more invasive and socially embedded applications.

The usefulness of the SEI could also be extended and deepened by having designer and participant look back over the video footage together, using it as reflective tool for discussing what was happening for the participant emotionally and why, and even brainstorming solutions and iterations together as a result of these insights.

4. Next steps

The results of this exploratory testing of the SEI lead us to believe that the instrument merits further refinement and more comprehensive testing.

One crucial next step is to determine the effectiveness of the SEI in working design contexts—are system developers able to use the SEI to elicit helpful feedback from users? What are the strengths and limitations of the tool in this active context? Does this vary across cultural locales/groups? We currently have research groups in both Sweden and Austria who are planning to use the objects during various stages of design, and we are soliciting designers in other countries—we are particularly interested in locating research groups in South America and Asia to generate a range of cultural responses to our instrument.

Based upon feedback from both US and Swedish participants in our study, we would like to make some changes to the object set. We hope to add objects that express the ‘missing’ range of emotions, to remove less helpful or redundant objects, and extend the overall shape range. We will probably keep the objects in the same size and color range, to limit the number of variables a person has to contend with, and to ensure that they stay comfortable and engaging to touch.

The patterns of response of both groups to the objects created by our sculptor suggested the following sculptural dimensions of interest:

- rounded vs. spiky (positive to negative valence),
- smooth vs. bubbly or protruding surface (low vs. high arousal),
- symmetrical vs. asymmetrical (calmer or more directed/resolved vs. confused/chaotic).

It is interesting to us that these three dimensions correspond to Osgood’s three ‘universal’ dimensions of subjective culture assessment cited in the introduction, and that the first two correspond to the two-dimensional valence and arousal space which forms the foundation for the International Affective Picture Set (IAPS) that was used in the calibration part of our study (Lang et al., 2005). Research into the workings of the brain suggest links between shapes and sounds (e.g. people will tend to map the name ‘kiki’ onto a sharp-edged drawing and ‘bouba’ onto a rounded-edged drawing—see Ramachandran and Hubbard, 2001), that seem to demonstrate fundamentally biological bases for our responses to visual properties. We see this work as a foundation for beginning to articulate ‘universal’ affective qualities of shapes, supporting their use in many cultural contexts. It is possible, though, that we may find after testing in radically different cultures that it is simply the case that different object sets are required—this would in turn have interesting implications for the limits of universality of emotion and expression of emotion in the core psychological literature.

As we make changes to the objects, we plan to explore iterations based upon the dimensions above (perhaps with the addition of the ‘droopiness’ dimension suggested by one participant). We may even create an interface that allows people to use sliders to craft their own affective objects, to give us a better understanding of these shape variables and how they relate to affect (the objects could then be ‘printed’ using a 3-d printing device and tested in further studies).

We also have a test arranged in which we compare self-report with the SEI, to measurements obtained through physiological tracking of the user—facial expression, skin conductance, heart rate, and pressure patterns on a pressure-sensitive mouse. It would also be interesting to compare results from the SEI to other self-report methods (such as PrEmo-style character-based ratings, think-aloud, and questionnaires). In future comparisons we also hope to include considerations of how participant age and gender (and other important demographic variables) may impact the effectiveness of the various methods.

Finally, after testing in working design contexts and iteration of the shapes, we would like to re-approach the question of cross-cultural validity of the SEI with a larger-scale test among cultures that are considered widely divergent. Our investigation of the literature on culture and design for this project has revealed a large knowledge

gap to us in the understanding of design process and designer/user communication as it varies among cultures. We see further study of the SEI as it functions in different cultural contexts as an avenue through which to approach this little understood yet essential area for designers of interfaces and products that will have global reach.

Acknowledgements

We thank the European Union for the funding of researchers dedicated to this project; we thank Rainey Straus for creating the objects, and for donating time to the project; and we thank those in the US and Sweden who participated in our studies. We thank Maria Cheadle for her assistance in analyzing the Swedish data. Thanks also to those affective presence researchers who inform and inspire our work: Phoebe Sengers, Bill Gaver, Michael Mateas, Geri Gay, Kirsten Boehner, Petra Sundström, Anna Ståhl, and Jofish Kaye.

References

- Averill, J.R., 1980. A constructivist view of emotion. In: Plutchik, R., Kellerman, H. (Eds.), *Emotion: Theory, Research, and Experience*. vol. 1: Theories of Emotion. Academic Press, New York, pp. 305–339.
- Bentley, T., Johnston, L., von Braggo, K., 2005. Evaluation Using Cued-Recall Debrief to Elicit Information about a User's Affective Experiences. In: *Proceedings of the OZCHI 2005*, Canberra, Australia.
- Berry, J.W., Poortinga, Y.H., Segall, M.H., Dasen, P.R., 1992. *Cross-Cultural Psychology: Research and Applications*. Cambridge University Press, Cambridge, UK.
- Bunch, B., 1989. *Reality's Mirror: Exploring the Mathematics of Symmetry*. Wiley, New York, NY.
- Desmet, P.M.A., 2004. Measuring emotions: development and application of an instrument to measure emotional responses to products. In: Blythe, M.A., Monk, A.F., Overbeeke, K., Wright, P.C. (Eds.), *Funology: From Usability to Enjoyment*. Springer, Berlin, pp. 111–123.
- Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., Scherer, K., Tomita, M., Tzavaras, A., 1987. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology* 53, 712–717.
- Gaver, W., Dunne, A., 1999. Projected Realities: Conceptual Design for Cultural Effect. In: *Proceedings of the CHI 1999*, Pittsburgh, PA, pp. 600–607.
- Isbister, K., Höök, K., 2005. Evaluating Affective Interfaces: Innovative Approaches. CHI 2005. http://www.sics.se/~kia/evaluating_affective_interfaces/
- Isbister, K., Höök, K., Sharp, M., Laaksolahti, J., 2006. The Sensual Evaluation Instrument: Developing an Affective Evaluation Tool. In: *Proceedings of the CHI 2006*, Montréal, Québec, Canada.
- Johnston, O., Thomas, F., 1995. *The Illusion of Life: Disney Animation*. Disney Editions.
- Kitayama, S., Markus, H.R., 1994. *Emotion and Culture: Empirical Studies of Mutual Influence*. American Psychological Association, Washington, DC.
- Lang, P.J., Bradley, M.M., Cuthbert, B.N., 2005. *International Affective Picture System (IAPS): Digitized Photographs, Instruction Manual and Affective Ratings*. Technical Report A-6. The Center for Research in Psychophysiology, University of Florida, Gainesville, FL.
- LeDoux, J.E., 1996. *The Emotional Brain*. Simon and Schuster, New York.
- Mandryk, R.L., Atkins, M.S., Inkpen, K.M., 2006. A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In: *Proceedings of the CHI 2006*, Montréal, Canada.
- Mesquita, B., Frijda, N.H., Scherer, K.R., 1997. Culture and emotion. In: second edBerry, J.W., Dasen, P.R., Saraswathi, T.S. (Eds.), *Basic Processes and Human Development*. vol. 2 of *Handbook of Cross-Cultural Psychology*. Allyn and Bacon, Boston, MA, pp. 255–297.
- Muller, M.J., Wildman, D.M., White, E.A., 1993. Taxonomy of PD Practices: A Brief Practitioner's Guide. In: *Participatory Design. Communications of the ACM (Special issue)* 36(4), 24–28.
- Myers, D.G., 2002. *Intuition: Its Powers and Perils*. Yale University Press.
- Norenzayan, A., Heine, S.J., 2005. Psychological universals: what are they and how can we know? *Psychological Bulletin* 131 (5), 763–784.
- Norman, D.A., 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, New York.
- O'Connor, K.J., 2000. *The Play Therapy Primer*. Wiley, New York.
- Osgood, C.E., 1977. Objective cross-national indicators of subjective culture. In: Poortinga, Y.H. (Ed.), *Basic Problems of Cross-Cultural Psychology*. Swets & Zeitlinger, Lisse, Netherlands, pp. 200–235.
- Picard, R., 1997. *Affective Computing*. MIT Press, Boston, MA.
- Picard, R., Daily, S.B., 2005. Evaluating affective interactions: alternatives to asking what users feel. Presented at *Evaluating Affective Interfaces: Innovative Approaches*. CHI 2005. http://www.sics.se/~kia/evaluating_affective_interfaces/
- Ramachandran, V.S., Hubbard, E.M., 2001. Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies* 8 (12), 3–34.
- Russell, J.A., 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (6), 1161–1178 (American Psychological Association).
- Russell, J.A., 1991. Culture and the categorization of emotions. *Psychological Bulletin* 110, 426–450.
- Russell, J.A., 1994. Is there universal recognition of emotion from facial expression? A review of cross-cultural studies. *Psychological Bulletin* 115, 102–141.
- Scherer, K.R., Wallbott, H.G., 1994. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology* 66 (2), 310–328.
- Wierzbicka, A., 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*. Oxford University Press, New York.
- Wierzbicka, A., 1999. *Emotions Across Languages and Cultures*. Cambridge University Press, Cambridge, UK.