INTRODUCING PREMO2
NEW DIRECTIONS FOR THE NON-VERBAL MEASUREMENT OF
EMOTION IN DESIGN

Gaël Laurans and Pieter Desmet,
Delft University of Technology.
G.F.G.Laurans@tudelft.nl, P.M.A.Desmet@tudelft.nl

ABSTRACT
Non-verbal emotion self-report methods have flourished in the last 10 years. We present the theoretical basis underlying non-verbal self-report, review available tools, and discuss perspectives for further progress in measurement practices in design-oriented emotion research. We describe a new questionnaire following these perspectives and the first steps of its validation. The results highlight the potential for improvements and limitations of existing measurement tools. The conclusion examines some lessons learned about the difficulties facing researchers interested in creating specific emotion measures.

Keywords: non-verbal emotion measure graphical questionnaire.

INTRODUCTION
One of the basic challenges in design & emotion research is to identify or develop appropriate procedures to reliably assess user experiences. Over the years, a wide variety of instruments has been used, both in academic and in practice-based design research, reflecting the multifaceted nature of user experiences. User experience is measured in, for example, concept testing of consumer products, interactive products, and services, and in research aimed at empathizing with user groups, in co-design and context mapping techniques. Some of the instruments used in the field are borrowed from psychology or anthropology, and some are developed (or customized) to fit particular characteristics of user experiences.

In academic research, carefully developed measures are needed to support quantitative research into the determinants of design-related emotions. In design practice, emotion measures can be used to assess new designs or compare different prototypes but also to inform the early phases of the design process and help define the user experience of a future product (Desmet & Schifferstein, 2012).

This paper first provides a brief review of emotion measurement tools, focusing in particular on pictorial scales, developed both in design research and in other domains. Building on this review, we discuss some limitations of the current tools and future perspectives of the maturation of the field. Next, a new instrument is introduced (PrEmo2). Preliminary validation studies in Korea and China and some ideas for future research are suggested, together with an overview of the lessons learned in the course of the development and validation of the new tool.

STATE OF THE ART
MEASUREMENT OF EMOTION
A large number of techniques have been applied to emotion measurement over the years, both inside and outside design research. They can be grouped according to the component of emotion into which they tap (Laurans, Desmet & Hekkert, 2009; Mauss & Robinson, 2009; Scherer, 2005). Thus psychophysiological techniques measure bodily changes accompanying emotional responses; facial expression recognition software uses expressive behavior associated with emotions and self-report questionnaires rely on the conscious feelings that form a crucial part of the emotional response. Of
these various techniques, self-report questionnaires remain the most common thanks to their versatility, ease of use and higher ability to discriminate between subtly different emotions (Laurans, Desmet & Hekkert, 2009).

Many questionnaires have been developed for the measurement of emotion and several of them have been used in a design context. Available questionnaires differ in response format, in underlying model (limited set of categorical emotions like fear and anger versus broad dimensions like positive-negative) and in the specific set of affect included (see Laurans, 2011; Laurans & Desmet, 2008 and Laurans, Desmet & Hekkert, 2009 for reviews).

By and large, traditional emotion questionnaires are focused on clinically relevant negative affect like depression and anxiety whereas positive emotions like interest, pleasure and fascination are typically lumped together in an undifferentiated “positive affect” or “joy” category. This limitation has long been noted by psychologists (Lorr & Wunderlich, 1988; Zuckerman & Lubin, 1990; Zuckerman, Lubin & Rinck, 1983) but it is particularly serious for design research as designers are especially interested in creating products that elicit positive experiences and need measures that are sensitive to subtle pleasant affective responses.

The lack of self-report questionnaires appropriate for the measurement of emotion outside of clinical contexts and the differentiation between different types of positive emotions led to the development of several new measurement tools specifically conceived for applied research in design (Desmet, 2002) or food science (King & Meiselman, 2010; Porcherot et al., 2010).

GRAPHICAL SELF-REPORT OF EMOTION
Self-report measures of affect can also be divided in two main groups according to the way emotions are represented. In verbal questionnaires, emotions are represented by words (emotion names like “fear” or adjectives like “afraid”) and respondents report their current state by indicating which words best match their feelings. Non-verbal questionnaires are based on the same principle, but instead of words emotions are represented graphically, for example through smiling or frowning faces.

While they can seem somewhat more complex to develop, graphical self-report instruments are particularly appealing for the measurement of emotional experience. Since they do not require research participants to verbalize their emotions, they do not rely on users’ being used to share and articulate their feelings with words.

Instead, they rely on the well-documented ability to attribute emotional meaning to non-verbal signals, in particular facial expression. Indeed, there is a large body of evidence showing that human beings can attribute some affective meaning to configurations of facial features and that these attributions are stable across cultures, at least partly (Ekman, 1992, 1999; Ekman, Sorenson & Friesen, 1969; Elfenbein & Ambady, 2002; Izard, 1971; Matsumoto & Willingham, 2006; but see also Russell, 1995; Russell, Bachorowski & Fernández-Dols, 2003).

Research on facial expression and non-verbal behavior in the 1960s to 1980s mostly focused on “basic emotions” (in particular Paul Ekman’s traditional list of six emotions: fear, anger, sadness, disgust, surprise, and enjoyment; Ekman, 1992) but more recent work also found evidence of broad recognition of other emotions like contempt and pride (Tracy & Robins, 2008) and emphasized the need to better discriminate between various types of positive affect (Sauter, 2010) a goal that is particularly relevant for design-oriented research.

Apart from photographs, simplified representations of facial movements (Bassili, 1978; 1979) and abstract moving shapes (Visch & Goudbeek, 2009) also carry affective information. Recent research also found dynamic displays of facial expression to elicit more activity in parts of the brain associated with the processing of emotional information than static facial expressions (Arsalidou, Morris & Taylor, 2011) and to induce clearer mimicry effects (Sato, Fujimura & Suzuki, 2008).

Voice parameters during speech have also been shown to convey affective information. Interestingly,
recent results suggest that abstract vocalizations expressing various basic emotions can be recognized across widely different cultures (Sauter, Eisner, Ekman & Scott, 2010).

Finally, graphical self-report have several practical advantages. As non-verbal representations do not depend on the emotion concepts embedded in a particular language, measurement instruments based on them can potentially be used in different cultures while eschewing translation of the emotion descriptors. This technique can also improve the validity of measures obtained from children or research participants interrogated in a second language or through a translator. While such approaches are understandably frowned upon in academic research they can still provide valuable insights in a design project.

These results provide a strong basis for the graphical representation of emotion in self-report questionnaires and suggest that this approach need not be restricted to “basic” negative emotions.

**REVIEW OF AVAILABLE INSTRUMENTS**

Before sketching some perspectives for the future of graphical self-report of emotion, we will review available tools.

The graphic differential (French, 1977) probably represents the oldest the type of graphical measurement of affective meaning. It is also atypical among more recent scales in that it does not rely mainly on facial expressions. Instead, the questionnaire comprises 15 bipolar “semantic differential” items, each anchored by pairs of pictures extracted from previous research following a similar approach. These pairs of pictures include for example a downward and an upward arrow, a standing tree and a broken one or a small and a large disc. Like in Osgood’s (1971) original semantic differential, these bipolar items are used to compute a score on three scales (evaluation, potency, and activation) found to describe most of the variance in affective meaning.

Bradley & Lang’s (1994) **Self-Assessment Manikin** (SAM) is a mood self-report instrument based on the same model of affect. Each of the dimensions (renamed pleasure or valence, arousal and dominance) is measured by a single item. The SAM uses schematic human figures to represent different feelings or affective states. It has been used in studies of intelligent homes (Dormann, 2003) and advertisement, among others.

Pic-a-mood (Vastenburg, Romero, van Bel & Desmet, 2011) is another scale designed to measure the first two dimensions of the pleasure-arousal-dominance model. Participants have to pick one of nine figures representing various moods defined by their position in the pleasure-arousal space. It is available in three forms (with a male, a female and a robot character) and has been used in studies of airport experiences, interaction with personal navigation devices or social media.

Schubert’s (1999) **2-Dimensional Emotion Scale** (2DES) is another similar scale used to collect moment-to-moment ratings of music experience. It comes as a software package that can be used to continuously record the emotion expressed by a musical excerpt while listening to it. Moving the mouse cursor in a bi-dimensional space anchored by basic smileys allows participants to report their experience of the music.

Several mood assessment tool measuring distressing experiences have also been developed to support research and clinical practice oriented toward children suffering from mental disorders.

For example, the **Koala Fear Questionnaire** (Muris et al., 2003) measures fear of 31 different situations using a three-point format anchored by a koala expressing different levels of fear. Buchanan & Niven (2002, 2003) **Facial Image Scale** and the **Venham Picture Test** (Venham, 1979) are comparable tools used for the measurement of anxiety in medical contexts (e.g. dental hospital). These questionnaires all rely on facial expressions to collect self-report of a particular emotion from pre-school children that often lack the cognitive ability and vocabulary necessary to express their feelings. The **Pictured Feelings Instrument** extends this approach to include different feelings like sad/depressed, fear/shock and angry/mad represented using 26 line drawings.
Assessment Via Animated Characters (MAAC; Manassissi et al., 2009) is a particularly interesting tool as it uses professional animations (not only static drawings) to represent 16 affective states (relaxed, bored, exhausted, surprised, sad, guilty, ashamed, angry, irritable, jealous, scared, nervous, disgusted, happy, elated, and pleased) to diagnose children suffering from anxiety disorders.

However, such questionnaires are not restricted to children and clinical samples. Several measurement instruments use the same principle to assess product experience in adults. The original PrEmo questionnaire (henceforth called PrEmo1 to distinguish it from the newer version presented below) was developed to measure people’s responses to product appearance using animated cartoons (Desmet, 2002, 2004). It measures between 10 and 14 individual emotions with a three-point response format. The LEMTool (Huisman & Van Hout, 2008) is designed to be overlaid directly on a website. Users can select one or more emotions from a list including joy, desire, fascination, satisfaction, sadness, disgust, boredom, and dissatisfaction. They can then rate the intensity with which they experience each of the selected emotions.

**FUTURE PERSPECTIVES**

From a user research point of view, many of the most widespread graphical questionnaires suffer from serious limitations. First of all, like regular verbal questionnaires, these tools have often been developed with clinical applications in mind and focus on the strong or repeated negative affective responses found in (young) anxious or depressed patients.

These scales are also quite rudimentary from a graphical standpoint and apart from PrEmo1 and the MAAC do not use the full potential of the various markers of emotion (posture, face, head and body movement) identified in the literature. Consequently a tool like the SAM, while often presented as a non-verbal tool is difficult to understand on its own and is actually meant to be used with extensive instructions explaining the meaning of the scales (Bradley & Lang, 2007). This reliance on (verbal) instructions begs the question of their translation in different languages and of the validity of the tool as a purely graphical self-report instrument. Questionnaires like PrEmo1 and the LEMTool do not suffer from this limitation but the content of these tools was defined with quite narrow applications in mind. Furthermore the validation data for nearly all tools mentioned in the review is limited to one or at most three or four countries and often based on small student samples.

For example, PrEmo1 development was informed by a series of studies providing an empirical basis for the design of the tool (Desmet, 2002, pp. 19-73 and 217-231). These studies were however based on a limited sample size and predated much interesting work in the field of design and emotion. Thus a comparison between the Geneva Emotion Wheel, a newly developed verbal self-report instrument (Scherer, 2005), and PrEmo1 (Laurans & Desmet, 2008) revealed that the latter lacked coverage of social or self-conscious emotions like pride, which have recently been the focus of much psychological research (Tracy & Robins, 2007a; Williams & DeSteno, 2009).

The initial PrEmo1 validation studies also suggested that participants had difficulties recognizing some of the emotion portrayed, which led to the redesign of some animations (Desmet, 2002, pp. 70-73), without fully resolving the problem.

As the field matures, it seems therefore increasingly important to develop tools with a more comprehensive theoretical basis and strong empirical validation.

**DEVELOPMENT OF PREMO2**

**GENERAL DESCRIPTION**

PrEmo2 is essentially an updated version of the PrEmo1 questionnaire and therefore a graphical emotion self-report tool comparable to the LEMTool or MAAC. Research participants (potential users or consumers) are asked to consider a list of emotions represented by animated cartoons and, for each emotion, to indicate the extent to which it corresponds to their current experience, using a five-point scale going from "I do not feel this" to "I do feel this strongly". The questionnaire can be administered
through a web interface (Figure 1), currently available in 15 languages. Participants have to provide a rating for each emotion before moving on to the next screen. They also have the possibility to correct their ratings before continuing to the next product. In a typical study, several products or several product pictures are presented to the participants at once before collecting ratings for each product one-by-one.

As in PrEmo1, the LEMTool or the MAAC, each emotion is represented by a cartoon character displaying a posture and facial expression typical of this emotion. Animation and sound further increase the expressivity of the cartoon.

**MAIN INNOVATIONS**

**Character style**
The most obvious difference between the new instrument and previously available tools is the character style, designed to maximize expressivity. The new characters are drawn from the waist up, using a frame similar to the LEMTool but narrower than that used in PrEmo1 or the SAM. This new frame gives more space to the face and upper body of the characters and makes for a more detailed representation of facial features. It is therefore much easier to discern the movement of the eyebrows, a major component of facial expression (Bassili, 1979; Eibl-Eibesfeldt, 1997, pp. 614-665), thus heightening the expressivity of the character (Figure 2).

Unlike existing “face scales” the new animations do however also include the upper body and the arms of the characters. Recent research highlighted the importance of these parts of the body to express “self-conscious” emotions like pride, shame and embarrassment (Tracy & Robins, 2007b; Tracy, Robins & Schriber, 2009). Finally, the new animations are colored to represent blushing (Figure 3).

**Emotion Set**
A somewhat less obvious but no less important difference is the set of emotions measured by the new questionnaire. As noted earlier, existing tools tend to target limited applications (product appearance for PrEmo1, interaction with websites for the LEMTool). For example, PrEmo1’s emotion set was constructed by reducing a large list of emotion names by collecting...
design students’ ratings of their relevance to the appearance of products – seeing them but not using or owning them (Desmet, 2002). Although this procedure can help identify product-relevant emotions neglected by psychological research, it lacks a sound theoretical basis.

![Figure 2. Stills from PrEmo1 (below) and PrEmo2 (above). From left to right: disgust, neutral position, and fascination.](image)

In the current work, a more systematic approach was adopted to devise a list of emotion that would be appropriate for a general product experience measurement tool. The model of emotion proposed by Ortony, Clore, and Collins (1988) was used to define four relevant emotional domains, further divided in 14 distinct emotions. These domains are (1) general well-being emotions (2) expectation-based emotions (3) social context emotions, and (4) material context emotions. Each domain is briefly described below.

To keep the questionnaire short and maintain the same number of positive and negative emotions, several emotions present in PrEmo1 are not included in PrEmo2: inspiration, indignation, disappointment, pleasant and unpleasant surprise. Table 1 provides a comparison of the emotions measured by PrEmo1 and PrEmo2. All emotions from the LEMTool are however covered by the new tool.

![Figure 3. Two stills from the “shame” animation with blushing, head and eyes lowered, shoulders slumped forward.](image)

(1) General well-being emotions
Joy or happiness is experienced when things that one likes happen. Conversely, sadness is experienced when things that one does not like happen, or when things do not happen that one wants. Emotions are also experienced in response the possible actualization of desirable or undesirable events. In those cases the actual prospect is unknown, and one will experience either hope or fear: Hope in the case one anticipates a desirable event to happen, and fear in the case one anticipates an undesirable event to happen.

(2) Expectation-based emotions
Expectations-based emotions are experienced in response to the conformation or disconfirmation of an expected event. The expectation serves a ‘standard of performance’. If the actual event confirms an expectation, a person will be satisfied and not merely happy. If the actual event disconfirms the expectation, a person will be dissatisfied and not merely sad.

(3) Social-context emotions
Emotions like admiration and contempt are experienced when events are attributed to the behavior of people or objects while pride and shame arise in responses to events we attribute to our own behavior. Admiration is evoked by behavior that is appraised as praiseworthy (exceeding a standard of behavior), and contempt is evoked by behavior that is appraised as blameworthy (conflicting with a standard of behavior).

(4) Material-context emotions
Attraction is experienced for objects that are appraised as desirable, and aversion for objects that are appraised as undesirable. Interest is experienced in response to objects that are appraised as distinctive
and boredom in response to objects that are appraised as unexceptional.

<table>
<thead>
<tr>
<th>PrEmo1</th>
<th>PrEmo2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasant surprise</td>
<td></td>
</tr>
<tr>
<td>Inspiration</td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>Fascination</td>
<td>Fascination</td>
</tr>
<tr>
<td>Amusement</td>
<td>Joy</td>
</tr>
<tr>
<td>Admiration</td>
<td>Admiration</td>
</tr>
<tr>
<td>Desire</td>
<td>Attraction</td>
</tr>
<tr>
<td></td>
<td>Pride</td>
</tr>
<tr>
<td></td>
<td>Hope</td>
</tr>
<tr>
<td>Unpleasant surprise</td>
<td></td>
</tr>
<tr>
<td>Indignation</td>
<td></td>
</tr>
<tr>
<td>Disappointment</td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td>Disgust</td>
</tr>
<tr>
<td>Contempt</td>
<td>Contempt</td>
</tr>
<tr>
<td>Dissatisfaction</td>
<td>Dissatisfaction</td>
</tr>
<tr>
<td>Boredom</td>
<td>Boredom</td>
</tr>
<tr>
<td></td>
<td>Shame</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
</tr>
<tr>
<td></td>
<td>Sadness</td>
</tr>
</tbody>
</table>

Table 1. Comparison of the set of emotions in PrEmo1 and PrEmo2.

**DEVELOPMENT PROCESS**

The development of the animations was inspired by the procedure described in Desmet (2002). Professional actors were filmed while enacting the set of 14 emotions. Six actors participated in total (one man and five women, all of them Dutch). All of them had extensive international acting experience, working for international theatre companies. Each actor or actress went through the procedure individually, and they did not see each other’s performance.

First, they were handed the emotion list, and explained the task, namely to express each emotion in few seconds using face, body, and vocalizations but without using actual words. They were allowed as much time as needed to prepare and practice the expressions. Once ready, the actors performed the task in a neutral room with a table (110 cm. high) in the middle of the room. A coffee cup was placed on the table and covered with a handkerchief. The camera was located in front of the actors, behind the table. The researcher showed the actors a sheet of paper with an emotion word. They had to nod to indicate when they were ready. Then, the researcher pulled the handkerchief from the coffee cup, which was the signal to initiate the expression. This procedure was designed to give the actors some physical reference to enable a natural emotional expression that is ‘about’ something. Moreover, pulling away the handkerchief enabled them to create an expression with a clear and sudden start (which is required for the animations).

This procedure was repeated for all 14 emotions. Each expression was recorded in one take, unless the actors were not satisfied with their performance. In that case new takes were recorded until they were satisfied. To facilitate the analysis, separate takes for the same emotion were grouped together. One researcher then discussed the general patterns emerging from the expressions of the different actors with the cartoon animator.

Finally, a distinct vocalization was recorded for each animation by a seventh actor not involved in the previous steps of the creation and synchronized to the animations.

**CHARACTER RECOGNITION**

This procedure allowed us to create compelling, purely non-verbal representations of the target emotions. However, the mere fact that these representations do not use words from a particular language does not automatically guarantee that the questionnaire will be totally culture-independent. No matter how careful the development process, it is still necessary to establish that research participants from various backgrounds interpret them similarly.
The first step of the validation is therefore a series of recognition studies examining the animations themselves, without any product rating or actual emotion self-report. These studies follow the general methodology employed in facial expression research and other judgment studies of non-verbal behavior (Wagner, 1997). The animations are shown one-by-one and the participants are invited to select a descriptor (emotion name) for each of them. This approach is not entirely satisfying because it requires a translation of the emotion descriptors and reintroduces language in the development of the questionnaire whereas one of the insights underlying this research is that it is difficult to verbalize one’s feelings and to translate emotion names across languages. Poor results in a recognition study could therefore stem from unclear or faulty animations but also from incorrect or ambiguous translation of the descriptors. Trying to match animations and emotion names is nonetheless useful to investigate the level of agreement between participants regarding the meaning of the animations and establish a broad correspondence between emotions and their representations.

The two studies reported here were conducted in Korea in 2007 and in China in 2008. These locales were chosen based on serendipitous research opportunities but also because interest for user experience and design research is growing in these countries and few if any measurement tools in the local languages can be found in the international literature.

Furthermore, cross-cultural studies of emotion recognition have found an “ingroup advantage” in the decoding of facial expression, where participants are generally better at interpreting facial expressions when the expressor is closer to them culturally and geographically (Elfenbein & Ambady, 2002, 2003a; but see Matsumoto, Olide & Willigham, 2009). The distance between cultures can be operationalized using Hofstede’s dimensions of cultural differences (Elfenbein & Ambady, 2003b). By that measure, Korea and China are two countries with a large cultural distance from the Netherlands and offer an ideal opportunity to test animations developed in this country. Recognition studies with PrEmo1, a previous version of the tool also developed in the Netherlands, found that some animations were more difficult to identify for participants in Japan and Finland than for participants in the Netherlands or the US (Desmet, 2002, p. 70), supporting this idea.

KOREA STUDY

Procedure
Participants (N = 18) were master students at a major technological university. The research was conducted in English. The animations were presented with a projector to all the participants at once and they recorded their answers on a response sheet. Because of this procedure, the order of the animations was fixed across all participants.

The response sheet used a forced choice format with 14 emotion names (“desire”, “satisfaction”, “pride”, “hope”, “joy”, “fascination”, “admiration”, “disgust”, “dissatisfaction”, “shame”, “fear”, “sadness”, “boredom”, and “contempt”) with no explicit “don’t know” choice. Additionally, participants were asked to indicate how sure of their choice they were, on a 7-point scale from “not sure” to “very sure”.

Results
The overall hit rate (proportion of correct answers across all animations and participants) was 76%. Recognition was excellent for most emotions, with hit rates between 70% and 100% for 10 of the 14 animations (Table 2).

Two animations, dissatisfaction and contempt, had serious problems and were often confused for each other, with “contempt” chosen by 56% of the participants for the dissatisfaction animation and “dissatisfaction” chosen by 44% of the participants for the contempt animation (another 28% chose “boredom”). Save for those two, negative emotions tended to be better recognized than positive emotions.

Two animations representing positive emotions, admiration and satisfaction, also had slightly lower hit rates even if more than half of the participants in the sample correctly identified them. Given the limited sample size, these observed hit rates are associated with a large error margin. Consequently, binomial
Confidence intervals using Wilson’s method (Agresti, 2002) were computed using the `binconf` function in R’s `Hmisc` package (Harrell, 2009). Observed hit rate was 61% for admiration (95% CI: [39%, 80%]) and 67% for satisfaction (95% CI: [44%, 84%]).

<table>
<thead>
<tr>
<th>Animation</th>
<th>Hit rate</th>
<th>Other choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attraction</td>
<td>83%</td>
<td>Admiration</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>67%</td>
<td>Pride</td>
</tr>
<tr>
<td>Pride</td>
<td>83%</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>Hope</td>
<td>72%</td>
<td>Admiration</td>
</tr>
<tr>
<td>Joy</td>
<td>72%</td>
<td>Fascination</td>
</tr>
<tr>
<td>Fascination</td>
<td>78%</td>
<td>Attraction</td>
</tr>
<tr>
<td>Admiration</td>
<td>61%</td>
<td>Fascination</td>
</tr>
<tr>
<td>Disgust</td>
<td>100%</td>
<td>NA</td>
</tr>
<tr>
<td>Dissatisfaction</td>
<td>33%</td>
<td>Contempt</td>
</tr>
<tr>
<td>Shame</td>
<td>100%</td>
<td>NA</td>
</tr>
<tr>
<td>Fear</td>
<td>94%</td>
<td>Disgust</td>
</tr>
<tr>
<td>Sadness</td>
<td>94%</td>
<td>Boredom</td>
</tr>
<tr>
<td>Boredom</td>
<td>100%</td>
<td>NA</td>
</tr>
<tr>
<td>Contempt</td>
<td>28%</td>
<td>Dissatisfaction</td>
</tr>
</tbody>
</table>

Table 2. Results from the Korea recognition study. “Hit rate” is the percentage of correct responses for each animation. “Other choice” is the most common incorrect response.

CHINA STUDY

The second recognition study was carried out in China in 2008. Beside a larger sample size, it improved on the Korean study in several ways: direct comparison with PrEmo1, translation of the emotion names, use of “don’t know” choice (Russell, 1994; Frank & Stennett, 2001), randomization of the animation and label orders, and broader demographics.

Procedure

The study took place in two locations in China (Shanghai and Chengdu), with 411 women between the age of 18 and 55 (M = 33, SD = 11) recruited by local research agencies. The whole procedure was conducted on-screen through a purpose-built web interface at the agencies’ facilities. Participants had to categorize both PrEmo1 and PrEmo2 animations. Each animation was presented in turn and had to be clicked and watched twice before answering the questions. Participants were first asked whether they thought the animation represented a positive or a negative emotion. Next, participants were asked to choose a single descriptor from a list of 14. For PrEmo1 animations, the original list of emotion names was used (“desire”, “satisfaction”, “admiration”, “inspiration”, “fascination”, “pleasant surprise”, “amusement”, “disgust”, “dissatisfaction”, “disappointment”, “contempt”, “boredom”, “indignation”, and “unpleasant surprise”). For the new set of animations, the list was “attraction”, “satisfaction”, “pride”, “hope”, “joy”, “fascination”, “admiration”, “disgust”, “dissatisfaction”, “shame”, “fear”, “sadness”, “boredom”, and “contempt”. For both questions, it was possible to select “I don’t know” and proceed to the next question. The orders of the set (PrEmo1 and PrEmo2), animations (within the set), and emotion names were all randomized across participants (but the order of the labels remained identical throughout each set for any given participant).

The study was conducted in Mandarin. All texts and emotion words were translated from English to Mandarin by a certified translator. The interpreter was briefed with background information on the context of the study (i.e. that the words describe emotions that people can experience in response to seeing, using, or owning consumer products). This procedure was adopted because previous studies involving the translations of emotion words revealed that they are heavily context-dependent (see e.g. Desmet, 2002). Finally, translations were back-translated by a second translator, and issues that emerged were resolved in discussions with the two translators.

Results

The overall hit rate was markedly lower than in the Korean sample, with 52% of correct identification across all animations and all participants. Hit rate for individual animations varied between 15 and 89% (Table 2). “Don’t know” choices represented between 0 and 9% of the answers, depending on the animation. Overall, the valence was however well
recognized with 94% correct across all valid answers (i.e. excluding about 3% “don’t know” answers).

Recognition was generally lower for animations representing positive emotions, with several of them being correctly recognized by less than a fifth of the participants. Despite these difficulties, the confusions were mostly limited to emotions of the same valence (i.e. participants choose the name of another positive emotion but did not confuse animations representing positive and negative emotions with each other), except in one case: hope.

“Hope” was the first choice for the corresponding animation but it was selected by only 22% (95% CI: [19%, 27%]) of the sample. Other frequent choices were “attraction” (15%), “fascination” (13%), “fear” (12%), “dissatisfaction” (8%), and “don’t know” (7%).

<table>
<thead>
<tr>
<th>Animation</th>
<th>Hit rate</th>
<th>Other choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attraction</td>
<td>15%</td>
<td>Hope</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>60%</td>
<td>Pride</td>
</tr>
<tr>
<td>Pride</td>
<td>54%</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>Hope</td>
<td>22%</td>
<td>Attraction</td>
</tr>
<tr>
<td>Joy</td>
<td>59%</td>
<td>Attraction</td>
</tr>
<tr>
<td>Fascination</td>
<td>15%</td>
<td>Attraction</td>
</tr>
<tr>
<td>Admiration</td>
<td>16%</td>
<td>Hope</td>
</tr>
<tr>
<td>Disgust</td>
<td>81%</td>
<td>Boredom</td>
</tr>
<tr>
<td>Dissatisfaction</td>
<td>59%</td>
<td>Contempt</td>
</tr>
<tr>
<td>Shame</td>
<td>69%</td>
<td>Sadness</td>
</tr>
<tr>
<td>Fear</td>
<td>73%</td>
<td>Dissatisfaction</td>
</tr>
<tr>
<td>Sadness</td>
<td>89%</td>
<td>Shame</td>
</tr>
<tr>
<td>Boredom</td>
<td>67%</td>
<td>Dissatisfaction</td>
</tr>
<tr>
<td>Contempt</td>
<td>66%</td>
<td>Dissatisfaction</td>
</tr>
</tbody>
</table>

Table 3. Results from the China recognition study. “Hit rate” is the percentage of correct responses for each animation. “Other choice” is the most common incorrect response.

Three other animations – attraction, fascination, and admiration – also had very poor performance. In all three cases the correct label was chosen by circa 15% of the participants and was not even the most common choice. The attraction animation was confused for “hope” by 46% of the sample, whereas the fascination animation was classified as “attraction” by 40% of the participants. Answers for “admiration” are spread out between “hope” (20%), “attraction” (19%), “fascination” (17%), and “joy” (10%).

Recognition of the PrEmo1 animations was lower across the board (overall hit rate 44%). Many emotions were recognized by less than half the sample and even animations representing negative emotions had hit rate between 37 and 63% (with one exception, unpleasant surprised, recognized by 84%).

Nearly all emotions represented in both sets were better recognized in PrEmo2 than in PrEmo1, sometimes dramatically so. Thus, the new disgust animation was correctly classified by 81% of the participants (95% CI: [77%, 85%]) compared to 55% for the PrEmo1 version (95% CI: [50%, 60%]), dissatisfaction by 59% (95% CI: [54%, 63%]) compared to 45% (95% CI: [41%, 50%]), contempt by 66% (95% CI: [61%, 71%]) compared to 61% (95% CI: [56%, 66%]), boredom by 67% (95% CI: [63%, 72%]) compared to 63% (95% CI: [58%, 67%]), amusement/joy by 59% (95% CI: [54%, 64%]) compared to 54% (95% CI: [49%, 59%]), and satisfaction by 60% (95% CI: [55%, 64%]) compared to 53% (95% CI: [48%, 58%]). Even the low-performing animation for admiration represents a clear improvement, with 16% hit rate for the new animation (95% CI: [12%, 19%]) against only 8% (95% CI: [5%, 11%]) for the original version.

The only major exception to this pattern of results is fascination. The fascination animation from the PrEmo1 questionnaire was recognized by 25% of the participants in this study (95% CI: [21%, 29%]) compared to only 15% (95% CI: [12%, 19%]) for the PrEmo2 version.

**GENERAL DISCUSSION**

These two studies tested the recognition of animations to be used in a new version of the PrEmo non-verbal self-report questionnaire. The recognition rates across the two samples provided encouraging results, particularly for negative emotions. While
positive emotions were generally recognized as such, discriminating between distinct positive emotions appears to be more difficult, especially outside of student populations.

The patterns of confusion do however suggest that participants were indeed able to distinguish more than one form of positive affect. For example, satisfaction and pride were often confused with each other but seldom with any other emotion. Thus, in both samples the most common incorrect answer for the satisfaction animation was “pride” and the most common incorrect answer for the pride animation was “satisfaction”. This result suggests that the two animations are interpreted in a similar ways independently of the culture of the respondents or the language used in the study. An intriguing hypothesis is that they represent somewhat related feelings or emotions that are only imperfectly captured by the verbal labels normally used to describe them.

Still, this interpretation does not change the fact that some animations seemed difficult to interpret for many participants, with answers spread across many categories. This is in particular the case for the hope (four distinct emotion names chosen by at least 10% of the participants) and admiration animations (five emotion names chosen by at least 10% of the participants).

CONCLUSION

The work presented in this paper represents the first validation effort of this scale for a design-oriented emotion measurement tool and one of the few to collect data beyond student samples in the design and emotion field.

Overall, the recognition results were mixed. The data confirmed that it is possible to represent emotions across cultures using animations. Importantly, these animations were not only very successful in conveying valence but also clearly allowed participants to discriminate different type of negative and positive affect, albeit not as many as hoped for.

The new animations also proved to be more easily recognized than corresponding animations representing the same emotions in the PrEmo1 questionnaire and several of the new emotions included in PrEmo2 gathered a high level of agreement. Nonetheless, some animations performed rather poorly in the general population Chinese sample. These results highlight both the potential for improvement in existing pictorial tools and the importance of systematic validation efforts going beyond student samples in one country.

Future research should therefore extend these tests to other countries and populations. Revision of the animations with the lowest performance could hopefully also alleviate some of the difficulties identify in this paper. Finally, using these animations to collect self-report of emotions in an actual product test is necessary to assess the reliability and validity of the questionnaire for design research.

ACKNOWLEDGEMENTS

Peter Wassink created the PrEmo2 character and animations, and Bob Kommer Studio created the sounds. We acknowledge David Güiza Caicedo for his valuable work on the development of the instrument and the actors who worked on the dynamic expressions. We also thank Chris Loxley, Lars Rengersen, Marco van Hout, and Woodear Wang in their support in organizing and running the China validation study. This research was supported by the MAGW VIDI grant number 452-10-011 of The Netherlands Organization for Scientific Research (N.W.O.) awarded to P.M.A. Desmet.

REFERENCES


