RELIABILITY OF EMOTIONAL RESPONSES TO MATERIAL TEXTURES

K. H. Hilton

Keywords: design, emotion, reliability

1. Introduction

Emotions are generally accepted to be a short-term implicit response to an experience, as opposed to a mood which is often a longer-term experience. As such, emotions may be experienced as initial responses which may fade quickly. A person may move through a series of emotions relating to a situation or article, before the experience settles for a period on an emotion or mix of emotions. These emotions might then be evoked again upon memory or re-experience of the article.

In order that we may more effectively design and market materials and products with emotional consideration, there is a need to gain a deeper understanding of the product-user relationship. This requires understanding to be built up from fundamental enquiry of response to decontextualised stimuli initially, determining how basic influences build upon and interrelate with one another to provide their complete sensorial experiences. This investigation used a range of domestic, workshop and clothing materials, see figure 1, in conjunction with Desmet's [2002] V5 PrEmo emotional caricatures, see figure 2, as an accessible example of facial expressions, for the elicitation of emotional responses. The aim was not to critically assess the materials or the caricatures but to consider the general reliability of human emotions, because it was questioned by the researcher 'just how reliable are emotions as a consideration within the design process?'

The method was designed to take two repeat-test approaches to gauging reliability of emotional experiences. The first experiment investigated the relationship of one stimuli set with another, in this case the textures with the caricatures, while the second experiment observed the placing of stimuli in positive-to-negative order of emotion. These two experiments were repeated a week later and the responses compared, to consider with which materials or caricatures reliability of response occurred. Initially the experiment was set to test participants blind in tactile mode only, in order to focus upon the sense of touch. However, because of responses and comments, it was considered important to run visual mode, and visual + tactile mode versions of the test, to see how the senses might be influencing reliability of emotional response.

This investigation considered under what circumstances emotional responses may be more reliable, by:

- a comparative ordering process.
- expression of associations through agreed reference coding.
- expression of associations through free verbal response.
- physical sense.
- the individual's subject skill.
- gender.
- declared mood.
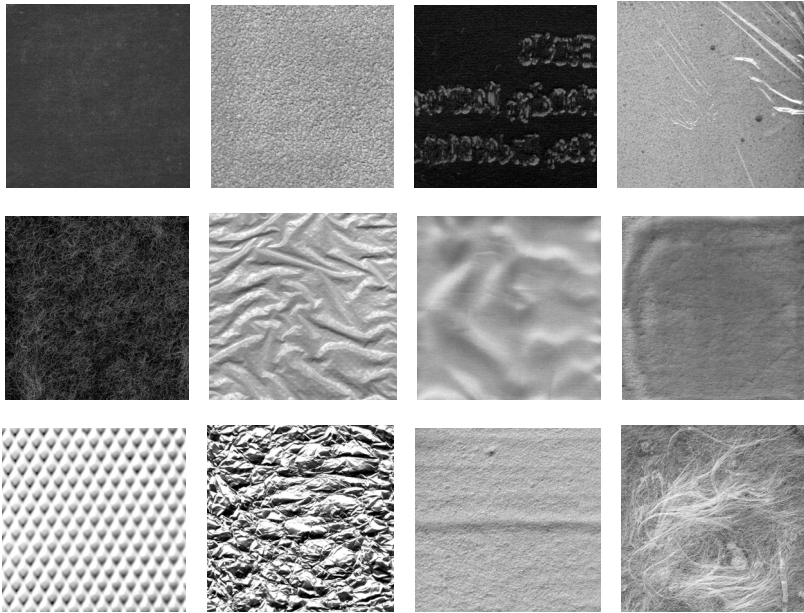- delivery order of the test.

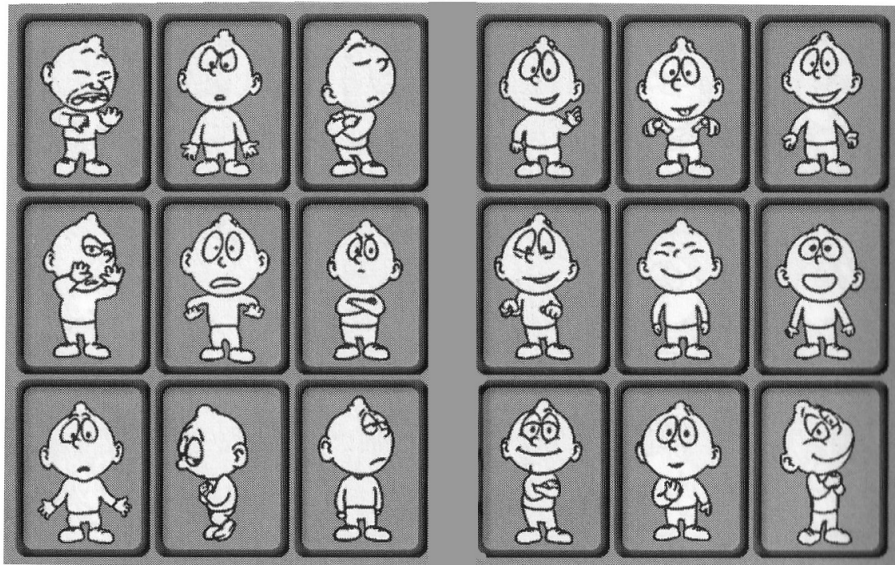**Figure 1. Texture sample grid. (Samples 1, 3 top left, 4, 1 bottom right.)**

Figure 2. The grid layout used of Desmet's PrEmo caricatures

2. Method

The test involved 50 participants, of whom 28 were males, 22 were females. The sample group composed a broad mixture of teaching, research, support staff, and students. The aim of this

investigation was to test the reliability of response to a range of 12 textures. See Figure 1. The experiments were run in one direction one week and then in reverse the following week to see whether participants would make the same connections between, and associations with, textures and caricatures, rather than make repeat responses by memory of the process. For example, when presented with a texture, a participant would select a caricature which they related emotionally to that texture, then the following week they were presented with that caricature to observe whether they selected the previously related texture. This experiment design was intended to discourage reliance or reference to memory, and encourage relationship with the materials.

It should be noted that the materials had colour, though illustrated here in black and white, and for the purpose of this investigation, the caricatures were used in what Desmet [2002] refers to as 'lazy' mode. The PrEmo tool is normally used as an interactive tool, which manages the test procedure, and includes sound and animation of emotions. However, for the requirements of this experiment the selected 18 caricatures were presented as images on paper only, used either in grid from or as separated images, simply as a resource of facial expressions.

An attempt was made to decontextualize the material samples, to make them less familiar, more of a new experience. This was achieved by placing them all on sponge backing in sample trays. This also aided delivery of the test, which may have proven problematic in tactile mode, had the samples been loose, because the grid layout of the texture set provided the coding method for the samples.

12 textures were chosen, as opposed to 18 in the grid of caricatures, see Figure 2, to avoid spatial connections being made between sample sets, and to avoid creating any sense of one correct association per texture.

The method was described to the participants prior to each test, but to avoid confusion each test was described to participants as a separate experiment. The wording of the tests varied in relation to differences in mode requirements. The following example was used for the combined visual + tactile mode:

> 'This experiment is in two parts. Test one and two take place today and test three and four take place in a week. After completion of the whole experiment there will be a debriefing to explain more about the investigation and its findings. However, it is possible to explain at this point that the purpose of the experiment is not to evaluate individuals.

> Test one requires that you familiarize yourself with a grid of 12 textures by looking at them closely and feeling them. Consider what each makes you think of and how they make you feel, emotionally. Then, when you are ready you will be shown a series of caricatures of emotion. You may pick a texture more than once throughout the test, and for each caricature you will be asked to declare the grid reference of the texture by column from the left, and then height from bottom. This declaration is for the texture which you believe is 'emotionally' most appropriate to the caricature shown. You may record 'none' if no texture feels appropriate. Then you may pass comment on your experience before going on to the next caricature. This test involves 18 expressions and is estimated to take 15 to 20 minutes.'

> 'Test two requires that you place the separate caricatures in a line arranged from most positive emotion on the left, to most negative on the right. Then place the separately boxed textures in a line, similarly from most positive to most negative.'

> (One week later)

> 'Test three requires that you familiarize yourself with a grid of 18 caricatures of emotion. Consider what each makes you think of and how they make you feel emotionally. When you are ready, you will be presented with a series of textures to feel and view inside separate boxes. Indicate which of the caricatures of emotion shown on the sheet provided is most fitting to each texture in turn. You may pick a caricature more than once throughout the test, and you should give the grid reference

letter first, A-F, number second, 1-3. After the grid reference you may pass comment on your experience before going on to the next texture. This test involves 12 textures and is estimated to take 15 to 20 minutes.'

'Test four requires that you place the separately boxed textures in a line arranged from most positive emotion on the left, to most negative on the right. Then place the separate caricatures in a line, similarly from most positive to most negative.'

In the case of the tactile mode, participants were tested with the textures concealed in a large box with sleeve-vents, for tests one and three. For the texture ordering in tests two and four they were required to wear a blindfold, an approach used effectively by Sonneveld [2002] for a contrasting textures concept project. In the case of the visual mode, where touch was not allowed, high definition prints, at the same scale, were substituted for the single and set boxed textures.

An additional methodological consideration, raised by Matravers [2001], was that emotional response may be influenced if the stimuli are not experienced on their own. The fact that under some of the test conditions these textures would be responded to as part of a set, it was accepted that this may influence the emotional impact registered. However, it was proposed that since people interact with everyday textures as part of the wider environment, the proposed test format was more ecologically correct. It is not possible to isolate influences completely anyway, only reduce influences to a level of fundamentality.

3. Results

Out of 12 textures and 18 caricatures, respondents were commonly only able to match 1 association in their second week. There were some however, who reselected between 3 and 5 in their second week. While this was not a significant percentage in terms of reliability (R), it was considered worth further checking. Of the many influencing factors on individuals, it was wondered whether dominance in sensory thinking style might be playing a part. For example low kinaesthetic thinkers might be expected to match none, whereas high kinaesthetic thinkers might be able to match a number of selections. However, it was decided that before possibly employing a suitable thinking styles test, a further reliability test should be run on the 'none' and the '3+' match participants. It was found was that the 'none' participants scored 'none' again, while the '3+' participants averaged 1 each. This indicated very low emotional response reliability for the primary stimuli, textures, associated with the secondary stimuli, the PrEmo caricatures in static 'lazy' mode. The sample codes were recorded as single figures, so for example, positions 3,2 and 4,2 were referred to as 32 and 43. See Table 1. This also avoided confusion with the numerical results shown in Tables 2 and 3.

In the analysis of the reliability of ordering emotional responses, there were seen to be a number of issues involved in considering the matches made. The approach taken established where the samples were placed most often in the linear order, and with what overall count the samples were placed, including the unmatched occurrences. Against this was a count for percentage of positional reliability, as there could be a number of different matches made at any particular position.

As anticipated, there was found to be greater comparability between the visual and visual + tactile mode responses, the visual sense being generally the more dominant and immediate mode of experience. In an examination of positional reliability it was indicated that whilst there can be no prediction that a given texture will be highly reliably placed, there was an indication of a greater reliability for those chosen for placement at the extremes.

During observation the researcher 'sensed' a trend in participant's placement of samples, which was not so clearly apparent in the statistical analysis. This was put down to attribution bias, where the researcher was possibly more aware of those sample arrangements which held closer to his own sense of positive to negative emotional order.

As emotions cannot be considered to have an exact nature, it was proposed that it would be unrealistic to expect exact positional matching on the repeat ordering exercise. So a level of flexibility, in a further analysis, was accommodated by awarding points to the level of 'vicinity' match to the previous week's ordering.

4 points were awarded for an exact match
2 for a match to a neighbour
1 for a match to a neighbour's neighbour
0 for anything more distant than a neighbour's neighbour

Table 1. Positional placement reliability of ordered samples

Tactile Mode	+											-
Sample Code.	32	42	42	43	32	12	22	33	31	21	21	23
Sample Count.	8	7	5	3	5	4	4	5	7	8	7	7
% Positional R.	40	40	0	20	20	0	20	10	10	40	30	60

Visual Mode	+											-
Sample Code.	13	12	21	22	43	23	31	11	43	41	41	41
Sample Count.	5	5	6	5	4	4	4	4	4	4	4	5
% Positional R.	30	10	30	20	20	10	30	30	30	20	20	40

Vis + Tactile	+											-
Sample Code.	13	13	42	32	33	22	43	11	21	22	12	41
Sample Count.	10	8	11	5	5	6	3	3	4	3	4	10

% Positional R. 60 50 60 50 30 20 10 20 30 20 20 70

When averaged across participants, this gave a more realistic sense of potential positional slew. As such, this analysis did not relate to a particular sample but the reliability of that position in the order. So for any person's order of samples, this would give a sense of whether reliability was relatively even across the samples, or whether there was greater likelihood of repetition at the extremes. See Table 2. The larger the number the greater the level of positional reliability.

Table 2. Vicinity match totals for textures

Tactile Mode	+ 2.1 2.2 1.3 1.6 2.0 1.5 2.1 1.5 1.7 2.7 2.1 2.8 -
Visual Mode	+ 2.4 1.8 2.1 1.9 1.6 0.7 2.1 1.7 1.7 1.9 2.0 2.2 -
Visual + Tactile Mode	+ 3.0 2.3 3.1 2.7 2.0 1.6 1.2 1.7 1.9 2.2 1.8 3.0 -

Again, there was an extremity slew, most apparent in the visual + tactile mode. The same approach taken with the PrEmo also indicated a slew. See Table 3.

Table 3. Reliability of response to PrEmo caricatures in 'lazy' mode

+ 2.3 2.7 1.9 2.0 1.8 1.2 2.2 2.0 1.9 1.9 2.1 1.7 1.9 1.5 1.5 2.5 2.8 3.3 -

The results were also checked as a group for modal response, to see if there were textures which were commonly referenced. This resulted in only a few 30-50% level modal responses across the sample, involving textures: 13, 23, 32 and 41. This modal response issue was checked in terms of gender to determine any differences in association tendency. It was found that males had a tendency to repeat reference to: 12, 21, 23, 32, and 41, while females had a tendency to refer to: 12, 22, 23, 31, 32, 33,

and 41. Further to this, it was found that with the emotional ordering of the textures, regardless of gender, some people's most positive textures were other peoples most negative. All the textures except 31 were positioned as most positive by at least one person and all textures except 32 were positioned as most negative by at least one person. However, it is proposed that with a larger group of participants these two textures would also likely be chosen as most positive and most negative by some people. The most positive texture overall was 13, the red velvet, while the most negative was 41, the matted brown synthetic fur. This showed that with large enough sample sizes it is possible to determine most and least favoured stimuli, but this would not translate to a guarantee of reliability of response of individuals.

A comparison of results for order of experiment delivery was also made. See Table 4.

Table 4. Comparison of delivery direction

	Average matching.	Range of matches.
Forwards delivery		
Re-associations.	2.0	0 - 5
Verbal connects.	3.8	0 - 9
Reverse delivery		
Re-associations.	2.0	1 - 3
Verbal connects.	1.5	0 - 6

Verbal association responses indicated differences in participant perception of focus, regardless of experiment instructions which set either the textures or the caricatures as focus depending on test mode. Participants were more able to repeat verbal associations with textures, than to associate textures with caricatures.

Participant's mood, at start of testing, was requested before each test to assess whether there was a connection between participant mood and any bias in the positive or negative sample ordering. The result was that there appeared to be negligible relationship for response to declared mood. It is possible that other factors like personality played a part in individual differences in the slews towards the emotional extremes, but this investigative method was not designed to analyze personality traits. The observation that emotional response can be somewhat detached of an individual's declared mood or frame of mind was not anticipated, and suggests an opportunity for further research as a fundamental issue in interaction experience. Another observation was that in some cases a negative slew on the ordering of the caricatures was accompanied by a positive slew on ordering of the textures. This suggested that either emotional response slewing may be subject or object dependent for some individuals, or random 'noise' may account for this observation. This might be investigated further by a longitudinal study with an extended sample group.

An additional observation made when tactile mode participants were finally allowed to see the textures they had been working with, was an experience of surprise with certain materials, an observation also made by Ludden *et al.* [2004]. Although participants were not all surprised to the same extent, or necessarily by the same samples, there was a tendency to acknowledge colours they had not expected, followed by material types which they had incorrectly identified, such as the foil and the cling-film. This error in identification was put down to the intended decontextualization by the foam backing of the samples.

4. Discussion

There is an acknowledged commercial interest within product design and marketing for a clearer understanding of emotions relating to product, with a view to managing and predicting change in consumers experience, [Hyun-Seung and Joohyeon, 2005]. Indeed interest includes the experience of change in a product as its surfaces wear, and how we relate to that ageing process. [Fisher, 2004].

Such investigative activities must remain clear in their working definitions, particularly regarding the difference between mood and emotion. Moods are longer-term states of mind, which make them eminently more predictable. Emotions however, are transitory and, as this research indicates in terms of attempts to reinitiate emotions in the following week, reliability is not as commonly experienced as might be expected. This poor predictability may be put down to individual differences and change, including how we are defining or imagining the emotional experience in question at the time. Mood was shown to effect self-reported emotions in the work of Forgas and Ciarrochi [2001]. However, little obvious effect was noted in this experiment. For example, some participants who declared themselves as unhappy had a positive emotion slew, others neutral and others negative. This is where personality may possibly influence how, and to what degree, mood can influence emotional response. It may be natural for some people when feeling low to compensate by making more upbeat reactions, or by being more aware of positive emotions from elsewhere.

Verbal responses collected overall, suggested that participants experiences were highly individual, relating to past instances and associations. The findings concerning slewing of extremes and poor reliability of subtleties might be extended to relate to other objects and suggestive of a potential lack of emotional reliability for 'things' in general, as opposed to 'situations' possibly. This has implications for the design profession who spend a significant amount of time working at the level of subtlety. 'Experience Designers' however, creating fairground rides and other forms of entertainment like computer games, may find they do tend to deal more with the extremes of emotion where response may prove more predictable, and where control of context may be much greater.

5. Conclusion

In the association of the primary, texture stimuli, with the secondary, caricatures, there was found to be a very low level of reliability of individual's emotional response. Even where participants were able to get a significant number of stimuli re-associated, a further test the following week for these high-scorers showed a more typical low re-association response. It is possible that if participants were told for both experiments that it was a memory test for the first week's emotional response, or a test of the participant determined choice of order, the reliability might actually have been greater, especially if the direction of the experiment was not reversed. The focus of this experiment however, had been on the emotional response felt towards each texture and caricature at the time of testing. The emotional ordering test indicated a degree of reliability slew towards the extremes, on both the textures and caricature tests. There was a greater slew for the textures in visual + tactile mode than the visual or tactile only modes. The analysis of modal responses across the sample group showed a number of textures were repeat referenced more successfully, but little more could be concluded than that these materials: velvet, sandpaper, satin and matted fur, may have been more readily identified with emotions. In terms of gender differentiation there was a sense from verbal responses that there was some difference in appreciation of certain clothing materials, where females more readily identified and described those samples.

The reliability of the test method was also assessed, but only direction of delivery seemed to create any difference to the connections made with verbal associations. This could have been down to the participant's interpretation of what the subject focus was, for the whole experiment. Mood was not determined to have a clear influence over the emotional response slews. Other individual differences possibly contributed to these slews, which may have bern subject dependent, as in some tests where the participants showed a slew for positive textures while showing a slew for negative caricatures.

The conclusion has been that the management of emotions within a consumer marketplace is questionable, and cannot be served well by subtleties. This suggests that much of the continued research on emotions will be of greater value to specific areas like 'Experience Design' and in the provision of certain design education experiences [Sonneveld, 2002], where the more extreme ends of the emotional spectrum may be given consideration. Success of inducing emotions in products generally, would seem to be dependent upon individual differences in experience, with their multitude of personal, interrelated influencing factors. Controversially, for emotion design research and practice, it is concluded that people's emotional experience of materials commonly lacks reliability. This suggests that effective application of emotion cues, as opposed to mood cues, in product design as a

whole may prove unreliable. It is expected that some researchers of design emotion may respond defensively to these findings, but that this paper should be seen to encourage further considerations and deeper discussion.

Acknowledgement

References

Desmet, P. "Designing Emotions." Delft University. 2002.

Fisher, T. H. "What We Touch, Touches Us: Materials, Affects and Affordances." Design Issues. Vol.20. Issue 4. 2004.

Forgas, J. P. and Ciarrochi, J. "On Being Happy and Possessive: The Interactive Effects of Mood and Personality on Consumer Judgments." Psychology and Marketing. Vol.18. Issue 3. 2001.

Hyun-Seung, C. and Joohyeon, L. "Development of a macroscopic model on recent fashion trends on the basis of consumer emotion." International Journal of Consumer Studies. Vol.29. Issue 1. 2005.

Ludden, G. D. S., Schifferstein, H. N. J. and Hekkert, P. "Surprises Elicited by Products Incorporating Visual – Tactual Incongruities." Conference proceedings for the 4th Design and Emotion, Ankara, Turkey. 2004.

Matravers, D. "Art and Emotion." Clarendon Press. 2001.

Plutchik, R. and Conte, H. R. (eds.) "Circumplex Models of Personality and Emotions." Washington: American Psychological Association. 1997.

Sonneveld, M. "Dreamy Hands: Exploring Tactile Aesthetics in Design." In McDonagh, D., Hekkert, P. Van Erp, J. and Gyi, D. (eds) Design and Emotion: Experience of Everyday Things. London: Taylor & Francis. 2002.

Dr. Kev Hilton
Head of Academic Research, School of Design
City Centre East Campus
Northumbria University
Newcastle upon Tyne
NE1 8ST
Tel.: 0191 243 7340
Email : k.hilton@northumbria.ac.uk
URL : http://www.openfolio.com/users/kevhilton

HUMAN BEHAVIOUR AND DESIGN