Comparing Perception of Affective Body Movements displayed by Actors and Animated Characters

Ariel Beck[1,2], Dr. Brett Stevens[1] and Dr. Kim A. Bard[3]

Abstract. This paper reports on a comparative study, which investigated how emotional body language, from animated characters and real actors, are perceived. The results are discussed in relation to the uncanny valley [1], which is a drop in believability as animated characters become more realistic [2]. The results showed that, videos of the actor were found to be more emotional, more believable and more natural than the animated characters whilst displaying the same emotional body language (recorded simultaneously by Motion Capture technology). Moreover, there was a significant difference in the number of correctly interpreted negative emotions displayed. Although, not for positive emotions. This could be due to the physical appearance of the animated character or to the loss of micro-gestures inherent to Motion Capture technology.

1 INTRODUCTION

Virtual environments (VE) for training have been shown to be efficient for skill learning [3, 4] by immersing the trainee in simulated situations. VEs allow participants to interact in a risk free environment. For example, in a medical scenario, a real patient is not required, reducing ethical and practical issues. Moreover, once developed, a VE can provide continuity across assessment contexts, which are then not subject to unpredictable variations arising from human intervention, as could be the case for training using actors. VEs also have many advantages when compared to traditional methods, such as video training, as they are responsive and flexible, adapting to instructions from the trainee in real time, whereas video training is always limited to a pre-recorded set of answers. However, it has been suggested that the key advantage of VE training is that it induces a feeling of presence (i.e. the feeling of being there), providing trainees with experience of situations that are more comparable to those evident in real life [5].

However, research on Animated characters populating the VEs used for the training of medical students has mostly focused on mechanical/surgical tasks. Nevertheless, work is beginning on simulating a wider set of patient interactions as well. This requires a full range of participant behaviours and actions, including speech, lip synchronisation, realistic looking anatomy and appearance, and animated facial expressions, all of which have started to be researched from a functional perspective [6]. However, recent results show the importance of body language as a medium for expressing emotions [7]. Hence, the animated character also requires more subtle behaviours such as non-verbal cues, or body language which would also be essential for training medical personnel in assessing a patient's inner emotional state.

To date, it is not known how users would interact with highly realistic characters and whether they would be considered as social partners. However, it has been shown that the way humans interact with technology is 'mindlessly' social [8]. It was found that technology can trigger social scripts, which typically apply to human-human interaction but are inappropriate for human-computer interaction, as they ignore the essential nature of the technology [8]. Moreover, consistent with this early result, recent studies support the fact that animated characters are indeed perceived as social agents, which trigger natural and social protocols in human users, such as gaze [9] or different acceptance levels based on ethnicity [10]. Nevertheless, these experiments considered mindless social rules and it is not evident that it will hold true at a conscious level.

The aim of the study presented, was to explore if an animated character would be consciously perceived as a social partner. This is done by testing the implicit assumption that viewers would perceive and interpret humans and animated characters in a similar way. Moreover, a similar approach as [8] can be adapted by testing the ability to perceive (i.e. see) and interpret (i.e. attribute meaning to) emotional expressions when displayed through the body [11]. This emotional body language is an ideal start to an investigation as it is known that people can accurately distinguish among emotions when displayed through human body language [12].

2 THE EXPERIMENT

2.1 Design

The main aim of the experiment was to compare emotional interpretation (accurate labelling and perceived strength) of a recorded person and an animation, which were both displaying the same emotional body language (Actor Vs Animated). However, in order to avoid the believability paradox, or Uncanny Valley [1], which is a drop in believability as characters acquire greater similarity with humans [2], traditional animation, such as Disney [13], usually uses stylised display of emotion, which are adapted to the physical appearance of the character. This could effect the perception of the animated display (Participants on seeing an animated character, could expect stylised and exaggerated display of emotions). Therefore the experiment also investigated whether there is an 'uncanny effect' in the display, by comparing the perceived believability and naturalness, of Stylised and Natural displays of emotion

[1] School of Creative Technologies, Univ. of Portsmouth, PO1 2DJ, UK
Email: {ariel.beck, brett.stevens}@port.ac.uk
[2] ExPERT Centre, Uni. Of Portsmouth, PO1 2FR, UK
[3] Centre for the Study of Emotions, Univ. of Portsmouth PO1 2DY, UK
Email: kim.bard@port.ac.uk

(Stylised vs. Natural). Studies on emotional body language recognition [7, 14, 15], consider only small sets of prototypical emotions, which could have resulted in discrimination between emotions (a process of elimination) rather than true recognition. This point is supported by Kramer (2008) who argues that nonverbal behaviours rarely carry specific intrinsic meanings but are interpreted within context [16]. Therefore, this study also investigates the effect of body language in two different situations, one with contextual verbal support and one unsupported (Voice vs. No Voice) on the interpretation of the Actor and Animated display.

Due to the believability paradox, which models a drop in believability as characters become highly realistic, the study investigates whether there is a relationship between perceived physical realism of the characters presented and the believability of the display.

For this study, it was necessary to consider the participants' ability to accurately interpret emotional behaviour, displayed by the actor and the animated character, and how this may affect the results. One of the main components of Emotional Intelligence is the perception of emotions [17]. Therefore its effects, including whether it correlates with the viewer's ability to classify emotion in either the video or the animated display, were investigated. Moreover, previous experience in video games and animated characters could also affect the perception of the animated character. Therefore, the experiments investigated possible correlations between previous experience and the results, including the number of good interpretations, perceived believability and naturalness.

2.2 Participants

25 Participants were recruited, mostly members of staff of the University of Portsmouth (16 females and 9 males) ranging in age from 24 to 61. Participants were randomly allocated to one of the two groups (Voice vs. No Voice). The thirteen participants in the Voice condition ranged in age from 24 to 61 (M=39.30, SD=13.70). The twelve participants in the no voice condition ranged in age from 31 to 60 (M=44.90, SD=9.66). Participants were entered in a raffle in exchange for participation.

2.3 Apparatus

A

B

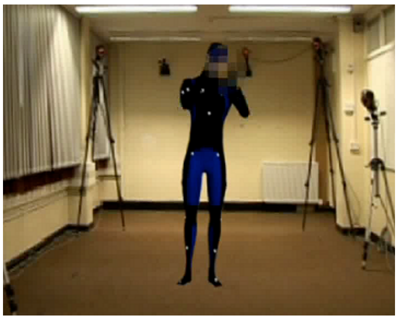

Figure 1. Screen shots of video condition (A) and animated condition (B).

A professional actor and a professional director were hired to build the material. The actor performed the following emotions: Anger, disgust, shame, fear, sadness, surprise, relief, happiness, pride and excitement. Each emotion was performed in two different ways, a natural version and an exaggerated stylised one closer to the one that can be seen in traditional animation. To ensure equivalency across conditions, the actor was video recorded (Figure 1 A) and motion captured simultaneously. The videos were recorded using a Sony PD170P. Motion capture data was recorded using a VICON motion capture system. The motion capture data was then used to animate a character (Figure 1 B) so that it displayed the same body language. The faces of the actor and the animated character were pixelated, so that this source of information was removed along with the possible uncanny effect that may come from facial animations. Moreover, to remove possible effects, such as differences in dress of the actor and animated character, both appeared in a motion capture suit (Figure 1), were physically similar (skin and face are not visible) and were put in the same context (Figure 1). Forty videos were created: 10 emotions x 2 versions x 2 types of display (see table 2).

These videos and animations were displayed on a 5m x 2.5m rear projection screen at life size. To record participants' answers, the material was embedded into custom made software, which was used for displaying the video clip as well as recording participants' answers.

2.4 Materials

The Emotional Intelligence (EI) was measured using the Emotional Intelligence Test-2nd Revision [18].
The physical realism of the actor and animated character were rated on a seven points Lickert scale using pictures showing them in the same 'neutral' position.

In order to record participants' interpretations of the emotional body language displayed, an existing questionnaire has been modified. It is based on the Geneva Emotional Wheel[1] that places twenty emotional labels on a two dimensional axis: valence and control. The Geneva Emotional Wheel is usually

[1] GEW; see [19], downloaded from http://www.unige.ch/fapse/emotion/ on the 01/07/2008

used for self-report (i.e. participants reporting their own emotional state). However, in this study, it has been used to report on participants' interpretations of the clips. The wheel's centre includes two additional options, "no emotion at all" and "none of the above", in case a different emotion is perceived by a participant. Each participant was also asked to indicate the strength for every emotional clip (five point Lickert scale radiating out from the centre).

For each emotional clip, participants were asked to rate the believability and naturalness on a seven point Lickert scale.
The Post-study questionnaire gathered personal information including gender, age, experience of video games and familiarity with animated characters. Finally participants were asked whether they noticed the presence of an animated character and if they found it disturbing.

2.5 Procedure

Participants were tested by the same experimenter in individual sessions. Each session began by obtaining consent, followed by an emotional intelligence questionnaire.

After completion of the questionnaire participants rated the realism of the actor and the animated character. It was necessary to rate realism before interpreting the video clips as if they watched the movements before, it would bias the rating. However, the fact that participants rated static pictures of the actor and animated character before interpreting the video clips may have affected the results. It can be argued that participants were made aware that it was a comparative study before watching the video clips and that this awareness could have biased their answers. However the experiment was advertised as a study on the recognition of emotional body language and the aim only explained during debrief. Moreover, several participants thought that they had been asked to interpret the exact same video clips more than once. This suggests that they were not aware of the comparative hypothesis during the test.

After rating the static pictures, participants started watching and assessing the video clips, which were played only one time. Once all video clips were interpreted, the post-study questionnaire automatically started. The whole procedure took one hour.

2.6 Results

Participants' mean score at the *Emotional Intelligence Questionnaire* was 111.92 out of 150 (SD=10.65). An average of 2.4 out of 7 (SD=1.47) was reported on *Experience in Video Games* and 3.8 out of 7 (SD=1.82) on *Familiarity with Animated Characters*. The mean rate of *physical realism* was 5.52 out of 7 (SD=1.83) for the actor and 2.76 out of 7 (SD=1.56) for the animated character.

Prior to any analysis, one-way ANOVAs have been computed to check if differences confounded the results between the two groups. There was no difference between groups in *EI* (F(1,23)=0.61, p=0.44), *Experience in Games* (F(1,18)=1.53, p=0.23) nor in *Familiarity with Animated Characters* (F(1,18)=1.53, p=0.47)

	Contextual Support	Presentation Condition	
		Acted Videos	**Animated Videos**
Believability		M=4.99 SD=1.01	M=4.38 SD=1.05
Naturalness		M=5.31 SD=0.99	M=4.51 SD=1.03
Perceived Strength		M=3.62 SD=0.63	M=3.33 SD=0.68
Number of Good Interpretations	**Voice** max=20	M=8.54 SD=1.94	M=8.08 SD=2.1
	No Voice max=20	M=5.83 SD=2.85	M=5.50 SD=2.15
Number of Good Interpretations (Negative)	**Voice** max=10	M=6.08 SD=1.32	M=5.08 SD=1.66
	No Voice max=10	M=3.25 SD=1.215	M=2.17 SD=1.115
Number of Good Interpretations (Positive)	**Voice** max=10	M=2.46 SD=1.39	M=3 SD=1
	No Voice max=10	M=2.58 SD=1.78	M=3.33 SD=1.67

Table 1. Means and Standart deviations of the results.

Repeated-measures ANOVA were carried out on the data. Assumptions of normality, homogeneity of variance and sphericity were met. Results showed that differences between conditions were unlikely to have arisen by sampling error. There was a significant difference in *Physical Realism* (F(1,23)=49.37, p<0.01), *Believability* (F(1,23)=12.4, p<0.01), *Naturalness* (F(1,23)=25.4, p<0.01) and *Strength* (F(1,23)=41.785, p<0.01). Repeated measures ANOVA did not show a significant difference in the *Number of Good Interpretations* (F(1,23)=0.9, p=0.4) for the acted and the animated condition, but there was a significant between subjects' effect for voice (F(1,23)=10.7, p<0.01). However, grouping the displays by valence revealed a significant difference for the negative ones (F(1,23)=14.5, p<0.01) for which participants were more accurate in interpreting the actor. There was no difference for the positive displays (F(1,23)=0.9, p=0.4).

To assess the correlation of *Emotional Intelligence* with the *Number of Good Interpretations*, Pearson correlations were computed. There was no correlation between *Emotional Intelligence* and the *Number of Good Interpretations* for the actor (r(23)=-0.3, p=0.09) nor for the animated character (r(23)=-0.2, p=0.1).

Similarly Pearson correlations were computed for *Emotional Intelligence* and *Believability*. There was no relation between *Emotional Intelligence* and the *Believability* of the actor (r(23)=0.27, p=0.10). However, *Emotional Intelligence* was positively correlated with the *Believability* of the animated character (r(23)=0.48, p<0.01). The results are similar with the perceived *Naturalness*, as there was no correlation between *Emotional Intelligence* and the *Naturalness* (r(23)=0.2, p=0.3) of the actor. However, *Emotional Intelligence* was positively correlated with the *Naturalness* of the animated character ($r(23)=0.5$, $p=0.01$).

Pearson's test showed a positive correlation between *Experience in Video Games* and the *Number of Good Interpretations* for the actor ($r(18)=0.37$, $p=0.05$) but not for the animated character ($r(18)=0.17$, $p=0.24$). However, *Familiarity with Animated Characters* was positively correlated with the *Number of Good Interpretations* for the actor ($r(18)=0.64$, $p<0.01$) as well as for the animation ($r(18)=0.38$, $p=0.05$).

There was a positive correlation between *Physical Realism of the Actor* with *Believability* of the actor ($r(23)=0.49$, $p<0.01$) and *Naturalness* of the actor ($r(23)=0.60$, $p<0.01$). However, there was no such correlation between *Physical Realism of the Animated Character* and *Believability* of the animated character ($r(23)=0.04$, $p=0.42$) nor with *naturalness* of the animated character ($r(23)=0.245$, $p=0.12$).

Finally, there was a positive correlation between *Experience in Games* and *Physical Realism of the Actor* ($r(23)=0.39$, $p=0.04$) as well as with *Physical Realism of the Animated Character* ($r(23)=0.75$, $p<0.01$). Similarly *Familiarity with Animated Characters* was positively correlated to *Physical Realism of the Actor* ($r(23)=045$, $p=0.02$) and to *Physical Realism of the Animated Character* ($r(23)=0.54$, $p<0.01$).

Emotion Displayed	**Condition**	**Good Answers in the Acted Condition**	**Good answers in the animated condition**
Anger	Voice	Natural: 13/13 Stylised: 10/13	Natural: 12/13 Stylised: 9/13
	No Voice	Natural: 3/12 Stylised: 10/12	Natural: 1/12 Stylised: 9/12
Disgust	Voice	Natural: 12/13 Stylised: 11/13	Natural: 13/13 Stylised: 9/13
	No Voice	Natural: 0 Stylised: 2/12	Natural: 2/12 Stylised: 1/12
Shame	Voice	Natural: 0 Stylised: 1/13	Natural: 1/13 Stylised: 3/13
	No Voice	Natural: 0 Stylised: 3/12	Natural: 0 Stylised: 0
Fear	Voice	Natural: 7/13 Stylised: 3/13	Natural: 6/13 Stylised: 2/13
	No Voice	Natural: 7/12 Stylised: 3/12	Natural: 4/12 Stylised: 1/12
Sadness	Voice	Natural: 11/13 Stylised: 9/13	Natural: 4/13 Stylised: 7/13
	No Voice	Natural: 4/12 Stylised: 7/12	Natural: 3/12 Stylised: 5/12
Surprise	Voice	Natural: 4/13 Stylised: 4/13	Natural: 3/13 Stylised: 0
	No Voice	Natural: 3/12 Stylised: 2/12	Natural: 2/12 Stylised: 1/12
Relief	Voice	Natural: 5/13 Stylised: 5/13	Natural: 6/13 Stylised: 9/13
	No Voice	Natural: 2/12 Stylised: 5/12	Natural: 5/12 Stylised: 6/12
Happiness	Voice	Natural: 1/13 Stylised: 2/13	Natural: 4/13 Stylised: 2/13
	No Voice	Natural: 1/12 Stylised: 4/12	Natural: 0/12 Stylised: 7/12
Pride	Voice	Natural: 0/13 Stylised: 3/13	Natural: 0/13 Stylised: 7/13
	No Voice	Natural: 0/12 Stylised: 4/12	Natural: 1/12 Stylised: 6/12
Excitement	Voice	Natural: 2/13 Stylised: 6/13	Natural: 0/13 Stylised: 8/13
	No Voice	Natural: 4/12 Stylised: 6/12	Natural: 5/12 Stylised: 7/12

Table 2. Proportion of good answers per emotions and conditions (13 participants in the Voice condition and 12 in the no voice condition). Chance level would be 1/22.

As shown in Table 2, recognition rates depended on emotion and condition, although most of the clips were recognised above chance level.

3 DISCUSSIONS

The results showed no difference in the number of video clips that were correctly interpreted between the actor and the animated character. However, it cannot be concluded that the interpretation of an actor and an animated character is similar, as the animated character has been perceived as less natural, less believable and less emotional. Moreover, grouping the displays by valence shows that for the negative emotions, participants were better at interpreting the actor. There was no difference for the positive emotions. It is important to emphasize, that these differences occurred with animated movements that were captured from the acted performance. However, four participants out of twenty-five reported not noticing the presence of animations, whilst only one participant reported being disturbed by the presence of animation.

The fact that no difference was found on the number of good interpretation may be due to a floor effect on the positive displays (i.e. Excitement, Happiness, Pride, Surprise, Relief): The recognition rate of these videos was very low in both conditions (See table 1). This may be due to the fact that it was impossible for participants to distinguish between so many positive emotional states watching emotional body language even with voice as well. However, this low recognition rate is consistent with Ekman's basic theory of emotion [20], which considers that happiness is at the core of all positive emotions. The basic theory of emotion suggests that the positive displays are very hard to accurately distinguish because of their similarities.

Motion Capture is not perfect; it fails to capture secondary animations such as skin deformation. This could explain why participants' interpreted more accurately the actor displaying a negative emotion than the animated character. This is shown by the significant difference for the video clips where a negative emotion (i.e. Anger, Fear, Disgust, Sadness, Shame) was expressed. Similar differences were found in the strength of the emotion, in believability and in naturalness where it was found that the actor was more emotional, more believable and more natural than the animated character. It can be suggested that these differences are due to the gestures that were not recreated by the Motion Capture Technology.

However, the differences in the physical appearance of the actor and the animated character could also explain the significant drop in believability and naturalness (Figure 1). The highly significant difference between the *realism of the actor* and the *realism of the animated character* (which were rated using pictures) revealed that the animated character failed to reach the right side of the Uncanny Valley. Nevertheless, it can be argued that the animated character used for this study is at least as realistic as the one that would be used for medical interview training. These characters have to be animated in real time and thus would also be susceptible to the same drop in believability and naturalness.

Participants who scored high on the Emotional Intelligence questionnaire were less affected by the missing micro-gestures and the lack of realism of the animated character. This was shown by the positive correlation between emotional intelligence, believability and naturalness of the animation. However, the lack of a correlation between Emotional Intelligence and the number of good interpretation could indicate that either the questionnaire failed to assess participants' ability to accurately interpret emotional body language or that the videos presented lacked the appropriate emotional cues. The second possibility can be dismissed by the fact that for almost all emotional clips the recognition rate is above chance level (Table 2).

The positive correlation between familiarity with animated character and the number of good answers in both conditions is very encouraging for medical interview training. It shows that participants who reported being familiar with animated characters were more accurate in interpreting the video clips including the acted ones. However, it is not possible to conclude that these participants learned from previous experience how to interpret emotional body language. The positive correlations between experience in video games and familiarity with animated character with the physical realism of the actor and animated character could simply imply that participants' familiar with animated character were less affected by the appearance of the actor and the animated character which was different from every day displays: they wore the motion capture suit and their faces were pixelated. This could also explain why participants familiar with animation were more accurate. Therefore, it will be necessary to determine whether it is linked to the appearance and whether it is related to specific types of experience with animated characters (different type of animations, interaction…).

Nevertheless, these differences were not unexpected and it can be argued that animated characters cannot be, in terms of believability and naturalness, as good as videos of a real person. However, the results show that participants were able to accurately interpret many emotions displayed by the animated characters.

4 CONCLUSION

The results clearly established that emotional body language animated with motion capture is significantly different when compared to the video motion. It was not possible to clearly establish if this loss was due to the micro-gestures that Motion Capture failed to recreate or if it was due to the physical appearance of the animated character, which was poorly rated when directly compared to the actor.

The next research step will investigate the effect of changing the level of physical realism as well as the effect of reintroducing the secondary animations that were missing. It is expected that increasing the level of realism will affect the difference in believability, naturalness and will explore the real shape of the Uncanny Valley. Whereas, secondary animations should minimize the difference in the number of good interpretations as they should give complementary cues necessary to accurately interpret the display.

This line of research is important because the feeling of presence is crucial for medical training to be efficient. Therefore, in order to be useable animated characters need to overcome the believability paradox.

ACKNOWLEDGEMENTS

This project is funded by the ExPERT Centre at the University of Portsmouth as part of a PhD programme. The authors would like to thank Dr. Guy Van De Walle for his advices as well as the director Matt Ghrist and the actor Teo Ghil for their excellent work. Kim A. Bard was partially supported by FEELIX GROWING, an European Commission FP6 project, IST-045169.

REFERENCES

[1] M. Mori, "Bukimi no tani [The un-canny valley] ." *Energy,* vol. 7, pp. 33-35, 1970.
[2] H. Brenton, M. Gillies, D. Ballin, and D. Chatting, "The Uncanny Valley: does it exist and is it related to presence ?," *Presence-connect,* vol. http://www.presence-research.org/, April 2005 2005.
[3] R. Aylett, A. Paiva, S. Woods, L. Hall, and C. Zoll, "Expressive Characters in Anti-Bullying Education," *publication13280,* 2005.
[4] W. Swartout, J. Gratch, R. W. Hill, E. Hovy, S. Marsella, J. Rickel, and D. Traum, "Toward virtual humans," *AI Magazine,* vol. 27, pp. 96-108, Sum 2006.
[5] M. Slater, A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, N. Pistrang, and M. V. Sanchez-Vives, "A virtual repirse of the Stanley Milgram obedience experiments," *PLoS ONE,* vol. 1, December 2006 2006.
[6] V. Vinayagamoorthy, M. Gillies, A. Steed, E. Tanguy, X. Pan, C. Loscos, and M. Slater, "Building Expression into Virtual Characters " in *Eurographics 2006*, Vienna, 2006.
[7] B. de Gelder, "Towards the neurobiology of emotional body language," *Nature reviews Neuro-Science,* vol. 7, pp. 242-249, 2006.
[8] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers," *Journal of Social Issues,* vol. 56, pp. 81-103, 2000 2000.
[9] H. Prendinger, C. Ma, and M. Ishizuka, "Eyes movements as indices for the utility of life-like agents: a pilot study," *Interacting with Computers,* vol. 19, pp. 281-292, 2007 2007.
[10] J. Pratt A, K. Hauser, Z. Ugray, and O. Patterson, "Looking at human-computer interface design: Effects of ethnicity in computer agents," *Interacting with Computers,* vol. 19, pp. 512-523, 2007.

[11] S. Pichon, B. De Gelder, and J. Grèzes, "Emotional modulation of visual and motor areas by dynamic body expressions of anger," *Social neuroscience,* vol.?, p.?, ? 2007.
[12] J. V. den Stock, R. Righart, and B. de Gelder, "Body expressions influence recognition of emotions in the face and voice," *Emotion,* vol. 7, pp. 487-494, 2007.
[13] F. Thomas and O. Johnston, *The illusion of life*. New-York: Abbeville Press, 1995.
[14] J. V. den Stock, J. Grèzes, and B. de Gelder, "Human and animal sounds influence recognition of body language," *Brain Research,* vol. in press, 2008.
[15] A. P. Atkinson, W. H. Dittrich, A. J. Gemmell, and A. W. Young, "Emotion perception from dynamic and static body expressions in point-light and full-light displays," *Perception,* vol. 33, pp. 717-746, 2004.
[16] N. C. Kramer, "Nonverbal communication," in *Human behavior in military contexts*, J. J. Blascovitch and C. R. Hartel, Eds., 2007.
[17] J. W. Kalat and M. N. Shiota, *Emotion*. Belmont: Thomson Wadsworth, 2007.
[18] I. Jerabek, "Emotional Intelligence Test - 2nd Revision," PsychTests.com, 2001.
[19] K. R. Scherer, "What are emotions? And how can they me measured?," *Social Science Information,* vol. 44, pp. 693-727, 2005.
[20] P. Ekman, "An argument for basic emotions," *Cognition and Emotion,* vol. 6, pp. 169-200, 1992.