

Show Me or Tell Me: Designing Avatars for Feedback

MICHELLE SCOTT¹, LUCAS PEREIRA¹ AND IAN OAKLEY^{2*}

¹*Madeira Interactive Technologies Institute, University of Madeira, Caminho da Penteadá, 9020-105 Funchal, Madeira, Portugal*

²*School of Design and Human Engineering, Ulsan National Institute of Science and Technology, Ulsan 689-798, Republic of Korea*

**Corresponding author: ian.r.oakley@gmail.com*

Avatars can be employed as a motivational tool, for example, allowing non-verbal communication that can be close to human communication. We describe two lab studies where we presented participants with avatars that communicated verbally via text and visually via expressions. In the first study, participants rated five different categories of captions and corresponding avatars. Results showed that the most persuasive, consistent and trustworthy verbal feedback was given in a humanized form. The second study was an exhaustive forced choice experiment where participants chose the happiest avatar from a pair displayed. Results showed participants found visual avatars more expressive and easier to understand than their verbal counterparts, and that users respond differently when presented with negative or positive emotions. This paper contributes to a better understanding of how to design feedback for expressive avatars.

RESEARCH HIGHLIGHTS

- An exploration of multi-modal personifications (avatars) for motivational feedback
- Validation of terms for five different styles of verbal response in avatars
- Empirical results contrasting expressiveness of gestural and verbal feedback in avatars

Keywords: agent-based interaction, avatars, user studies

Editorial Board Member: Dianne Murray

Received 23 November 2012; Revised 13 February 2014; Accepted 19 February 2014

1. INTRODUCTION

People frequently behave in ways that harm or damage themselves, others or their environment; they smoke, make poor dietary and exercise choices and waste natural resources through excessive or inefficient consumption. Frequently they are aware of the negative impact of their actions (West, 2005). A substantial psychological literature addresses this space by seeking to understand, encourage and motivate behavior change, typically toward objectively positive outcomes such as smoking cessation (Curry *et al.*, 1991), increased exercise levels (Consolvo *et al.*, 2008) and reduced energy consumption (Abrahamse *et al.*, 2005). The potential impact of such research is substantial. Behavior change research has shown considerable potential in supporting people to consume less (Abrahamse *et al.*, 2005), and authors have argued that

psychological understandings are key to understanding why individuals engage in unsustainable consumption behaviors and in designing interventions to encourage more globally viable choices (Koger and Scott, 2007).

Fogg (2003) argues that computers can effectively persuade users in tasks as diverse as purchasing decisions (Dormann, 2000), diet management (Lee *et al.*, 2011) and energy consumption (Kimura and Nakajima, 2010). Techniques that computers enable include dynamic delivery of just-in-time notifications, highlighting of contextually relevant cues and display of rich, interactive, engaging visualizations such as avatars (Fogg, 2003). This paper focuses on design issues for one these tools: expressive graphical avatars. Technologically, such personalized visualizations are an ideal platform for motivation. They are everyday feature of today's digital media platforms, such as games (and game consoles) (Xbox LIVE

Avatars) and social networking services such as Second Life or Cyworld. Via modular third-party services such as Buddy Poke (www.buddypoke.com), avatars can be easily integrated into common tools such as websites. Creating a humanized website, including emotive text, pictures of people and virtual agents, has been shown to increase trust in e-commerce (Head *et al.*, 2003). Avatars are also frequently reported to be both expressive and engaging and have the potential to attract sustained long-term interest (Fabri *et al.*, 2002). Reflecting these advantages, authors have included avatars (or virtual pets (Dillahunt *et al.*, 2008; Lin *et al.*, 2006), a highly related concept) in behavior change scenarios in a range of domains including health (Lin *et al.*, 2006), energy consumption (Mahmud *et al.*, 2007; Dillahunt *et al.*, 2008) and education (Baker *et al.*, 2009). However, such systems are typically designed *ad hoc* and rely on simple metaphors, such as the growth of a garden of flowers (Consolvo *et al.*, 2008) or the abundance of natural animals in a scene (Lin *et al.*, 2006). This paper argues that avatars can be more effective motivators by adopting more powerful and evocative communication channels.

Despite the prevalence of avatar systems, there is relatively little literature relating to how they can be systematically designed in order to be motivating. In a recent review paper, Baylor (2011) provides an explanation for this by highlighting the complexity of designing avatars. She specifically identifies appropriately conveying emotional expressions to enhance motivation as a challenging design problem. Practical work on this topic has confirmed these notions. For example, in an educational scenario, Baylor and Kim (2009) found that avatar facial expressions effectively motivated learners, while deictic gestures negatively impacted students' attitudes. The authors suggest that difficulties in interpreting the gestures reduced the effectiveness of the communication. Similarly, Theng and Aung (2011) presented avatars with happy and sad facial expressions to children depending on their performance in a quiz. Although the emotional communication was reported to increase understanding, the authors found lower levels of motivation when the children were shown the sad faces.

This finding directly relates to the body of literature indicating that excessive negative feedback discourages behavior change (Deci, 1972). This literature states that behavior change must fundamentally be driven by motivation or desires, and makes a number of distinctions about the form that motivations can take. Perhaps most importantly it defines intrinsic motivation as a process of interest and enjoyment where the activity has inherent satisfaction for the person. In contrast, extrinsic motivations occur when the goal of the behavior is separable from the activity itself, either in the form of punishment avoidance or in the pursuit of a valued outcome. A person can transition from no motivation to extrinsic motivation to intrinsic motivation in a particular activity by internalizing and integrating the values and behavior regulations typical to that domain (Deci *et al.*, 1994). However, this process is not automatic, and rewards, such as money, food and prizes, have generally been found to undermine

intrinsic motivation (Deci *et al.*, 1999). Intrinsic motivation has been shown to have more powerful effects on long-term behavior change compared with extrinsic motivation (Curry *et al.*, 1991). Therefore, in order to have long-term effects on sustainable behavior, intrinsic motivation must be encouraged. Research indicates that both rewards and rebukes represent key aspects of a behavior change intervention (Deci *et al.*, 1999), but that they need to be treated differently—positive feedback generally encourages users, while overly negative responses can be demotivating (Deci, 1972).

This paper extends this work. Specifically, one notably omitted aspect is the impact of emotions expressed by virtual personifications via lexical content or facial and body language. This paper argues that such affective cues would be a beneficial component contributing to the persuasive power of personifications. This paper explicitly contrasts the presentation of such messages in an avatar via two modalities: verbal (textual messages) versus visual (expressions and body language). Prior to a full comparison of these modalities, a validation study is conducted to best match verbal captions with graphical emotions.

There are two main contributions in this paper. First, we validate a set of domain neutral images and textual content providing positive and negative feedback intended for use in avatars within behavior change scenarios. We believe this study to be novel within the domain of behavior change research and that the findings, in the form of validated visual and textual media, will be useful to future designers who wish to use avatars to present persuasive feedback. Secondly, we present the results of an experiment indicating where and when to deploy each of these message modalities in order to maximize the effectiveness of such a behavior change system. The aim of this work is to inform the design of avatars as a motivational tool by creating actionable guidelines that recommend how feedback should be deployed.

The remainder of this paper is structured as follows: a review of emotional communication in avatars, a description of two studies exploring textual and visual depictions of positive and negative feedback in the context of graphical avatars and a discussion of the results of this work.

2. EMOTIONAL COMMUNICATION

How avatars can best express their feelings and the impact this has on users has been the topic of extensive research. Many authors report the formation of strong bonds. Bailenson *et al.* (2001), for example, suggest that users respond to virtual avatars as if they were human, even if they are highly stylized and unrealistic. Boberg *et al.* (2008) report that some users go as far as seeking personal interaction and physical contact, such as hugs and touches, with virtual avatars, while Benyon and Mival (2008) argue that users can come to form attachments to avatars that are sufficiently strong that damage to their virtual

representations is perceived as damage to themselves. Other work has revealed the persona effect (Lester *et al.*, 1997), where a lifelike character can have a strong positive effect on learning. The authors compared different levels of expressiveness of an animated pedagogical agent, including animated advice, verbal advice and no advice regarding a task students had to complete. Interacting with the agents had a positive effect on the students' test scores; however, the mute agent and the task-specific verbal agent were not as effective as the more expressive agents. The authors argue that the presence of an animated agent has a positive effect on students' perception of the learning experience and that this effect could increase their motivation to learn. The persona effect has been empirically studied (Mulken *et al.*, 1998) and the authors found that the presence of a persona has a positive effect on participants' impression of a presentation. Participants perceived the presentations as more entertaining and less difficult, although there were no differences in comprehension and recall of the presentation compared with an absent persona. Taken together, these findings highlight the powerful effects of emotional communication shown via virtual personifications.

However, coherently explaining the quality and nature of these attachments is a current research challenge; conflicting accounts exist. Nass *et al.* (1994) conducted a series of experiments to determine whether interactions with computers are social. They found that participants had social reactions to computers and that these reactions were not dependent on the belief that the computer was human or human-like. Participants, despite being experienced computer users, reacted with politeness and responded socially to the computer itself. The authors argue that these social responses are automatic and unconscious and call this the "Computers as Social Actors" paradigm. This work has been extended to the Ethopoeia explanation (Nass and Moon, 2000) that suggests that interactions between humans and emotional virtual agents will unfold similarly due to humans unconsciously applying social rules to computers. Studies supporting this framework have provided evidence that people respond politely to computer agents, ethnically identify with them and apply gender stereotypes to them. In contrast, the Threshold Model of Social Influence (Blascovich, 2002) argues that social responses to virtual agents will take place only if the presentation of the agent is sufficiently realistic so that it cannot be distinguished from a human. However, the author states that some categories of behavior are more important than others, particularly that emotional expressions are more important than large limb movements. In recent work comparing these approaches, von der Pütten *et al.* (2010) examined social interaction with a virtual character that was said to be either computer or human controlled, and which had either low or high behavioral realism. The results indicated no differences in participant's behaviors, a finding that strongly supports the Ethopoeia model and its assertion that humans readily form emotional bonds to artificial characters.

Research has also focused on how to design emotionally expressive and understandable avatars. Numerous display channels are available including facial expressions, body language or animations and verbal or lexical messages; all can be effective, but the design of appropriate emotional cues remains a challenge. For example, Baylor (2011) states verbal communication should be as human as possible and that visual gestures need to be carefully designed. Krämer *et al.* (2007) demonstrate this point in their exploration of the impact of a range of non-verbal cues including self-touching (scratching), which led to positive impressions, due to the agent being rated as more natural and warm-hearted and eyebrow raising, which led to negative opinions and feelings. This situation becomes more complicated when cues are presented spanning multiple modalities—evidence indicates that incongruent or poorly matched pairings of cues reduce comprehension. Hong *et al.* (2002) illustrated this point in a study combining happy, neutral and sad voices with happy, neutral and sad faces. Their results indicated that it was more difficult for users to accurately identify emotions with incongruent cues. Similarly, Creed and Beale (2008) showed that mismatches between the emotions presented vocally and via facial expressions led to inconsistencies as study participants attempted to resolve the conflicting cues into a single coherent percept. This work shows that consistency in the delivery of multi-channel messages via avatars is key to understanding them. De Gelder and Vroomen (2000) present findings that further emphasize these difficulties, suggesting that incoherent expressions take longer to interpret and lead to more neutral ratings of overall emotion (e.g. a happy voice and a sad face will be rated as somewhere between these two emotions). Together this literature stresses the importance of consistent emotional cues to facilitate ease and speed of comprehension.

A smaller body of work has looked at purely graphical presentations in the same way. These typically combine printed text with pictorial avatar expressions. For instance, Carroll and Russell (1996) used a story to provide context to depicted facial expressions. Study participants were read a story, shown a photo of a face and then asked to choose an emotion for the face in a forced choice paradigm. The results indicated that contextual cues, in the form of the lexical narrative, exerted a strong influence on the perceived emotion. They conclude that facial expressions cannot be considered in isolation. More recently, Noël *et al.* (2009) described a study comparing emotions depicted in avatar facial expressions and short captions presented in speech bubbles. Five emotions were considered (anger, happiness, neutral, sadness and surprise) and congruent and incongruent presentations evaluated. Participants were able to identify emotions accurately independently of congruence, but facial expressions exerted a stronger influence on interpretation of the text.

Depicting a neutral emotion can cause difficulties in interpretation. For example, Noël *et al.* (2009) reported challenges in validating the cues they studied. The neutral avatar

face was frequently reported to depict a sad expression and the sad text had a low overall recognition rate (30.8%). Creed and Beale (2008) also showed that animated faces showing a neutral face were less engaging, warm and concerned than those showing a happy or warm face. Casanueva and Blake (2001) compared static avatars to expressive, gesturing avatars within a collaborative virtual environment and found the expressive avatars were rated as having higher co-presence. Co-presence is defined as the feeling that other participants within the virtual environment are actually present and feel like real people. Finally, Dyck *et al.* (2008) compared pictures of human faces with avatar faces on ratings of happiness, sadness, anger, fear, disgust and neutral. They found that neutral expressions were more easily recognized on a human face than on an avatar. They also found that for the avatar faces, sadness was confused with the neutral face, supporting similar findings by Noël *et al.* (2009).

In summary, the goal of this paper is to investigate how expressive avatars can be used as a motivational tool. It is motivated by the idea that virtual personifications are effective motivational tools (Fogg, 2003) and seeks to flesh out aspects of idea. Specifically, it suggests that previous work has focused on contextually descriptive visualizations such as the health of an environment (Consolvo *et al.*, 2008) or the abundance of natural resources (Dillahunt *et al.*, 2008) rather than leveraging emotionally expressive avatars as a persuasive tool. This focus reflects the complexity of displaying emotions in avatars, particularly if multiple modalities, such as textual messages and facial and bodily expressions are used (e.g. Baylor, 2011). Ultimately, this paper argues that expressive avatars utilizing multiple modes of communication have the potential to be a powerful, compelling motivational tool. The empirical work contributes to the design of motivational avatars by showing how they can combine expressions and textual context in order to be both easily understandable and effective persuaders. This work is described in the subsequent sections.

3. STIMULI VALIDATION STUDY

This study was conducted to select and validate emotionally expressive stimuli, in the form of textual captions and pictures of avatars, for future use. This experiment replicates some parts of previous work (Noël *et al.*, 2009). Their participants saw human and avatar faces (happy, sad, neutral, angry or surprised) accompanied by congruent or incongruent text. They were interested in incongruencies between text and expressions and their participants rated the faces (displayed below the text) on trustworthiness, sincerity, appropriateness, intensity and convincingness. In the current study, the emotional range used had five levels spanning very sad to very happy, including a neutral middle. These are three of the seven main emotions used in previous relevant work, the others being surprise, anger, fear and disgust (Fabri *et al.*, 2002). This dimension was selected because we were interested in a clear positive and negative range of emotion. The study presented participants with graphical avatars and captions and asked them to rate the individual cues in terms of their persuasiveness, happiness and trustworthiness and the pair of cues in terms of their consistency.

One set of graphical avatars was used, all of which are shown in Fig. 1. The graphical avatars were generated from existing templates on the BuddyPoke service (www.buddypoke.com). The five expressions used were selected according to categories on the BuddyPoke system and validated via subjective assessment by an experimenter. The particular avatar character, a manga-like cartoon human, was selected as it is gender-neutral in terms of both its facial appearance and its clothing. The choice of a gender-neutral avatar was to ensure we also used the same cue set while also minimizing the influence of participant gender on the experimental results. Items in five different categories of caption (goal-related, social, performance, humanized and comparison) were presented. These categories of caption were chosen to reflect some commonly used behavior change interventions, e.g. goal setting and feedback (Locke and



FIGURE 1. Full set of visual avatars with verbal humanized captions used in validation study.

TABLE 1. Full set of captions.

Category	Emotion				
	Very sad	Sad	Neutral	Happy	Very happy
Goal related	You didn't even come close to your goals.	You seem to have forgotten your goals.	You could set more goals.	Good, you achieved almost all your goals.	Congratulations on achieving all your goals!
Social	Everyone feels like you are failing them.	You're letting the team down.	Try to support your teammates.	Everyone is happy with your help.	You are really supporting everyone!
Performance	You are failing miserably.	You can do better work.	You seem to be on track.	You have done well.	Nice, you did a really great job!
Humanized	You've totally let me down.	I'm disappointed with you.	I want you to keep working.	I like your work so far.	I couldn't be more proud of you!
Comparison	Everyone is better than you.	You're doing somewhat worse than others.	You are doing similar to others.	You're starting to do better than others.	You've done much better than others!

How happy is the character on the right?

Very Unhappy Very Happy

1 2 3 4 5 6 7 8 9

“You’re doing somewhat worse than others.”

How happy is the previous caption message?

Very Unhappy Very Happy

1 2 3 4 5 6 7 8 9

How happy consistent is the caption message with the character:

Very Inconsistent Very Consistent


1 2 3 4 5 6 7 8 9

Rate the persuasion of the:

Not Persuasive Very Persuasive

Character 1 2 3 4 5 6 7 8 9

Message 1 2 3 4 5 6 7 8 9



Progress **Next**

FIGURE 2. Screenshot of validation study.

Latham, 2002), social pressure and comparison with others (Lin *et al.*, 2006) and also an attempt to utilize the tendency to treat virtual avatars as human (Bailenson *et al.*, 2001) These are presented in Table 1. A screenshot of the final experimental interface is presented in Fig. 2.

In each trial in the study, a graphical avatar expression was shown with one of the emotionally corresponding captions—the graphical and textual equivalent cues always showed equivalent emotions. The full complement of avatar and caption combinations were generated and presented in a random order.

TABLE 2. Mean ratings of happiness, trustworthiness and persuasiveness by the emotion level of the characters. Standard errors are shown in brackets. F values calculated from repeated-measures ANOVA on the mean ratings for each happiness level are also shown.

Rating scale	Emotion level: visual					Within-subjects ANOVA
	Very happy	Happy	Neutral	Sad	Very sad	
Happiness	8.31 (0.11)	7.12 (0.14)	4.89 (0.08)	2.10 (0.13)	1.64 (0.12)	$F(4,36) = 655.78, P < 0.01$
Trustworthiness	7.51 (0.21)	7.27 (0.19)	6.35 (0.23)	6.98 (0.20)	7.13 (0.17)	$F(4,36) = 14.25, P < 0.01$
Persuasiveness	7.56 (0.21)	7.26 (0.17)	6.15 (0.24)	7.18 (0.22)	7.27 (0.20)	$F(4,36) = 18.04, P < 0.01$

Each pair appeared twice in order to increase the reliability of the responses. This led to a total of 50 trials. The experiment took approximately 40 min to complete.

3.1. Participants

Twenty participants completed the study. Participants were recruited through an online advertisement via the university email list. The ages of the participants ranged from 21 to 33 with a mean age of 25.8. Six of the participants were female and 14 were male. Fourteen of the participants were educated to undergraduate level or higher, one had some college and five had completed high school. All of the participants were students. The majority of participants were Portuguese, 12 in total, while six were from the USA and the final two were from India and Kenya. All reported their English levels to be good or higher. Participants were compensated for completing the study with a 10 Euro voucher for a popular local retailer.

3.2. Procedure

The experiment was conducted at a computer terminal in a lab. The opening screen showed one of the avatars with a speech bubble and welcome message and detailed experimental instructions were available to view. Because of the self-explanatory nature of the experiment, viewing of the instructions was not enforced. This page also captured basic demographics: language, age, employment, nationality, education level and gender. After entering data, participants clicked a button to move to the experiment itself.

Each trial consisted of an avatar and caption being displayed on the screen. Participants were asked to make five ratings on 9-item Likert scales (see Fig. 2). These were labeled as avatar happiness and caption happiness (both scales ranging from very unhappy to very happy), the consistency of the avatar with the caption (from very inconsistent to very consistent), the persuasiveness of the avatar and caption (from not persuasive to very persuasive) and how trustworthy the avatar and caption appeared (from not trustworthy to very trustworthy).

After completing these ratings, participants were asked to answer some open-ended questions on paper. These included general thoughts, how expressive or irritating they felt the

combination of text and avatars was and how expressive or irritating they were separately.

This study was intended to be exploratory and conducted to select the most persuasive, consistent and trustworthy captions to use in future work. As such, we had no formal hypotheses as to which category of caption would be found to be the most persuasive, consistent and trustworthy.

4. RESULTS

4.1. Ratings

Table 2 shows the raw data and output of a set of within-subjects ANOVA tests run with the emotion levels of the expressions as the factors. The mean ratings of happiness in the table show clearly that participants rated the happiest expression as the happiest, the saddest as the saddest and so on. A within-subjects ANOVA confirmed a significant main effect of emotion level of the expression on ratings of happiness and pairwise *post hoc* comparisons [least significant difference (LSD)] showed the neutral emotion level to be significantly different than the other levels of emotion on ratings of happiness ($P < 0.01$). There was also a significant main effect of emotion level on ratings of persuasion and trustworthiness. In fact, significantly higher ratings of trust and persuasion were recorded for the very happy emotion ($P < 0.01$) compared with the other emotions shown by the avatar. Once again, *post hoc* comparisons showed the neutral emotion level to be significantly different from the other levels of emotion on ratings of persuasiveness and trustworthiness. As can be seen in the table, the neutral expressions were rated as the least persuasive and trustworthy.

Table 3 shows the raw data and output of a set of within-subjects ANOVA tests run with the emotion levels of the captions as the factors. It also shows the ratings for the degree of consistency reported between visual and verbal cues. Ratings of happiness showed a significant main effect: higher ratings were assigned to happier emotions, indicating the captions successfully communicated their intended messages. This trend was borne out by significance in the *post hoc* (LSD) tests between every happiness level ($P < 0.01$). There were significant main effects of caption category on ratings of consistency, trustworthiness and persuasiveness. Ratings of the trustworthiness and persuasiveness of the individual verbal

TABLE 3. Mean ratings of persuasiveness, consistency, trustworthiness and happiness by the emotion level of the captions. Standard errors are shown in brackets. F values calculated from repeated-measures ANOVA on the mean ratings for each happiness level are also shown.

Rating scale	Emotion level: verbal					Within-subjects ANOVA
	Very happy	Happy	Neutral	Sad	Very sad	
Happiness	7.87 (0.12)	6.73 (0.10)	4.76 (0.09)	2.77 (0.12)	1.66 (0.11)	$F(4, 36) = 583.9, P < 0.01$
Consistency	7.45 (0.22)	7.54 (0.15)	6.47 (0.24)	6.77 (0.21)	7.18 (0.21)	$F(4, 36) = 8.52, P < 0.01$
Trustworthiness	7.63 (0.17)	7.31 (0.17)	6.70 (0.20)	7.20 (0.19)	7.37 (0.21)	$F(4, 36) = 9.93, P < 0.01$
Persuasiveness	7.55 (0.16)	7.27 (0.16)	6.60 (0.19)	7.19 (0.16)	7.48 (0.18)	$F(4, 36) = 11.94, P < 0.01$

TABLE 4. Mean ratings of persuasiveness, consistency and trustworthiness by category caption. Standard Errors are shown in brackets. F values calculated from repeated measures ANOVA on the mean ratings for each happiness level are also shown.

Rating scale	Caption category					Within-subjects ANOVA
	Goal	Social	Performance	Humanized	Comparison	
Persuasiveness	6.96 (0.20)	7.23 (0.15)	7.51 (0.16)	7.56 (0.15)	6.80 (0.18)	$F(4, 36) = 9.95, P < 0.01$
Consistency	6.88 (0.19)	7.03 (0.19)	7.17 (0.17)	7.44 (0.19)	6.87 (0.17)	$F(4, 36) = 4.87, P < 0.01$
Trustworthiness	7.26 (0.19)	7.20 (0.16)	7.40 (0.19)	7.49 (0.17)	6.86 (0.19)	$F(4, 36) = 7.92, P < 0.01$

cues and consistency of the pair revealed significant trends. In all cases, these took the form of lower ratings for the neutral emotion, suggesting that, however they are instantiated, neutral statements lack trustworthiness and persuasiveness and, ultimately, are challenging to interpret; this finding was significant in *post hoc* tests ($P < 0.01$).

A similar set of within-subjects ANOVAs were run with category of caption as the factor. The means and output are shown in Table 4. There were significant main effects of category on ratings of persuasiveness, consistency and trustworthiness. As can be seen in Table 4, the humanized category was rated as the most persuasive, consistent and trustworthy. Pairwise comparisons (LSD) indicated that the humanized captions were rated significantly higher than all the other categories, apart from performance, on ratings of persuasiveness and consistency ($P < 0.05$). Due to the number of comparisons run on this data, the authors advise caution in interpreting these results significant at the 5% level. Pairwise comparisons also showed that the humanized category was significantly different from the comparison category only on ratings of trustworthiness ($P < 0.01$).

4.2. Qualitative measures

All 20 participants answered open-ended questions regarding the avatars and captions. The first main theme to emerge was that almost half of the participants wanted a greater range of expressions for the avatars. One stated: ‘the avatars were very limited in terms of expressions, there should have been more of them.’ The second theme that emerged was that the extreme emotions were easier to interpret. Users suggested that the avatars were better at expressing emotions at each end of the spectrum rather than neutral emotions: ‘avatars seemed very

expressive and would be appropriate for extremes but not for ‘you’re doing better than others’’. Another user stated: ‘the one that’s kind of in the middle . . . she doesn’t look happy or sad, just kind of in the middle looks slightly emotionless . . . it only really suited the ones where it said “you are exactly the same as others”’. This matches the quantitative finding that the neutral avatars were the lowest rated on persuasiveness, trustworthiness and consistency. A third theme was that the combination of captions and expressions was more meaningful than either of them separately. For instance, one remarked, ‘the text helps put a context to feelings.’ The final theme to emerge was that the potential for nuance and increased expressive weight was another reported benefit of the combined of message modalities. Once participant stated, ‘the avatar probably softens the blow of negative messages, and does the reverse for positive.’ Participants did not report the avatars and captions to be irritating or annoying.

5. DISCUSSION

The character and caption analyses showed that participants clearly rated the emotion levels of the captions and characters according to the five pre-defined emotions. This confirms the captions and characters effectively expressed an affective spectrum ranging from very sad to very happy. The mean happiness ratings for both character and caption at each emotion level were also numerically very similar reinforcing the consistency expressed between the captions and expressions. In the qualitative interviews, people commented that the combination of caption and visual expression was more meaningful and more contextually appropriate than having them displayed separately. Consistency of the cue pairs was also

explicitly measured. Of the five emotions studied, the neutral expression and caption were seen as the least consistent pair. This suggests that the neutral emotion is the least easy to understand—participants found it hard to relate the two neutral emotional expressions to one another.

Participants also found it hard to understand the neutral expressions by themselves—they were rated as significantly less trustworthy and persuasive than the other emotions for both the avatars and the text. In contrast, the more extreme emotions, particularly at the positive end of the scale, were rated as significantly more persuasive and trustworthy. The qualitative interviews also support this point, with participants stating that the stronger emotions were easier to understand. Perhaps unsurprisingly, this suggests that such low-intensity expressions of emotion are unlikely to be effective or convincing motivators.

Examining the data describing the five different caption categories, it is clear that all were effective: scores are close and consistent across all four of the attribute scales. However, of the four categories, the humanized category had optimal scores for persuasiveness, trustworthiness and consistency, suggesting that participants identified most strongly with an avatar that appeared to be addressing them personally and directly. In summary, this study suggests adopting a personal point of view in textual feedback accompanying graphical avatars will be easiest to understand and most effective. As suggested by previous work, a humanized design helps to engender trust (Head *et al.*, 2003).

6. EXPERIMENTAL DESIGN—COMPARISON STUDY

6.1. Overview

The goal of this study was to explore the impact of presenting positive and negative information to users via different communication channels, specifically emotional content expressed through graphical avatars and textual captions associated with them. The ultimate aim was to contrast how users perceive information in these two modalities in order to inform the design of avatars as a motivational tool. The study was based on the humanized category of captions, previously identified as the most persuasive and trustworthy as well as consistent with the graphical depictions of the avatars.

The study involved presenting pairs of avatars featuring either an emotional graphical expression or a neutral graphical expression plus emotional humanized caption. Participants were asked to choose which avatar they felt was happier. The full spectrum of five emotions (very happy to very sad) was presented in both modalities and the study followed an exhaustive forced choice design. Barring identical pairs, each possible combination of avatars was shown in a random order to each participant twice. In these pairs, the screen location of the avatar images was switched such that no images appeared consistently on the left or the right. The final screen layout can



FIGURE 3. Screen layout of comparison study.

be seen in Fig. 3. This design led to a total of 90 comparisons in each study. Prior to commencing the study, a practice set of 10 comparisons was displayed. These trials were generated randomly from the available set of combinations and there was no noticeable break between the practice and experimental trials. The experiment took no longer than 10 min to complete.

6.2. Participants

Forty-six participants completed the study. The ages of the participants ranged from 19 to 45, with a mean age of 28.9. Twenty-eight of the participants were male and 18 were female. The majority of participants were educated to undergraduate level or higher, 35 in total, 7 had some college education and 5 had completed high school. Twenty-eight participants were employed, 17 were students and 1 was unemployed. The majority of participants came from the UK, 19 in total, the rest were from 14 different countries. All reported their English levels to be good or higher. Participants were not compensated.

6.3. Procedure

The experiment was conducted online. All materials were in English. The opening screen showed one of the avatars with a speech bubble and welcome message and detailed experimental instructions were available (optionally) to view. A brief questionnaire captured demographics and verified that each participant completed the study once only.

Each trial in the study started with a blank screen showing a fixation spot. Participants were required to press the spacebar to move on. They were then presented with two avatars, one on the left and one on the right of the screen, and the question: ‘Which one is happier?’ A screenshot of the study can be seen in Fig. 3. The ‘F’ key was used to answer this question with the leftmost avatar and the ‘J’ key to select the rightmost avatar. The trial then ended and the fixation spot for the next trial appeared.

7. RESULTS AND DISCUSSION

7.1. Overall happiness

Rather than analyze binary data, analyses were conducted on the percentage of times each stimulus was selected as the

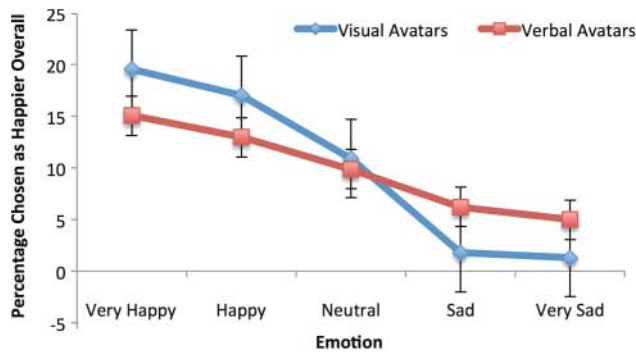


FIGURE 4. Comparison between avatar types on the overall percentage they were chosen as happiest.

happiest of the presented pair. These data are shown in Fig. 4. A two-way within-subjects ANOVA on the variables of emotion (very happy to very sad) and type (verbal and visual) revealed a significant main effect of emotion on happiness ratings ($F(4, 42) = 1705.40, P < 0.01$), but not of type ($F(1, 45) = 3.36, P = 0.73$). There was a significant interaction effect ($F(4, 1) = 138.43, P < 0.01$), also shown in Fig. 4. This can be interpreted as indicating that the visual avatars expressed the most extreme emotions—the number of ratings of both happiness and sadness in the visual case exceeded those in the verbal case.

7.2. Emotion level comparison

To compare the effectiveness of verbal and visual avatars at each specific emotion level, the percentage of times each stimulus was selected as happier than another expressing the same emotion level was calculated. These data are illustrated in Fig. 5. Separate within-subjects t -tests were then run to compare between avatars of the same emotional level in the verbal and visual conditions. These revealed significant differences at $P < 0.01$ for every emotion bar the neutral emotion. To ensure an acceptable family-wise error rate in this set of five t -tests, the critical value for significance was set at $P < 0.01$.

As can be seen in Fig. 5, there is a clear crossover in these data. The visual happy avatars were more frequently selected as happy when compared with the verbal happy avatars and conversely, the verbal sad avatars were more regularly rated as happier than the visual sad avatars. This suggests that avatars based on visual expressions are both more vivid and easier to interpret than those that use verbal expressions. This reinforces the point that participants viewed the visual avatars as more emotional—they expressed greater extremes of both happiness and sadness. It is also worth noting that, although this effect is lower in the happy end of the spectrum, it remains powerful when depicting sad emotions. This highlights the strength of the visual depictions of sadness and suggests that the unhappy avatars were highly effective at conveying their messages.

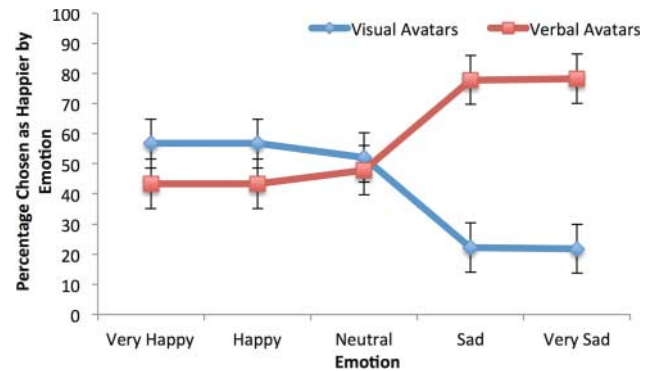


FIGURE 5. Comparison between avatar types at each emotion level; percentage chosen as happiest for each emotion is shown.

8. OVERALL DISCUSSION

These two studies explored how avatars should best present positive and negative feedback. The first study validated different categories of captions to accompany graphical avatars and was used to select the category of caption that was found to be most persuasive, trustworthy and consistent. The second study presented the graphical avatars and chosen category of caption in a forced choice paradigm to determine which channel of communication was perceived as happiest, visual or verbal. The two studies had conflicting objectives: the first tried to find consistency between the visual and verbal modes, whereas, in the second study, we attempted to tease apart the effects of these different channels.

The first study showed that humanized captions were the most persuasive, trustworthy and consistent. There were not large differences in the categories for the study, however, suggesting that the other categories may also be useful as persuasive feedback. In the second study, visual feedback was found to be more powerful than verbal feedback, the visual happy avatars were seen as happier and the visual sad avatars were seen as sadder than the verbal avatars. Study 1 also showed a similar pattern: the mean very happy and happy ratings of the characters were higher than the mean very happy and happy ratings of the captions. Study 1 also showed higher ratings of trust and persuasion for the very happy visual avatar compared with the very sad. In the qualitative data from Study 1, one participant stated, ‘the avatar probably softens the blow of negative messages, and does the reverse for positive.’ In general, the cartoon-style avatars also appear to be effective communicators of emotion, supporting the Ethopoeia theory (Nass and Moon, 2000) that suggests that humans respond to computers socially without the need for fully realistic, human-like presentations.

Participants in both studies gave very similar happiness ratings for the different levels of emotion in the visual avatars; they elicited highly consistent subjective assessments across the two studies. They also showed a higher recognition rate than

that reported in a comparable work (e.g. Noël *et al.*, 2009). This could be attributable to the choice of avatar. Noël *et al.* used a computer-generated face, which may have presented more subtle or ambiguous cues to people compared with the cartoon style deployed in the current work. Regardless, the neutral emotion was challenging to interpret in both studies, supporting previous work (Noël *et al.*, 2009; Creed and Beale, 2008; Dyck, 2008). It was least trustworthy and persuasive for both the captions and the character in Study 1 and there were no differences in Study 2 between verbal and visual for the neutral emotion. This suggests that, in an avatar feedback system, there may be no need to present neutral feedback. It may be clearer and less confusing for users to see only positive or negative feedback. However, it also may be true that abrupt changes from positive and negative presentations may be undesirable, potentially confusing or annoying users. We identify managing the transition from positive to negative feedback, ideally by avoiding ambiguous neutral content, as an area that deserves further study.

Previous work on using avatars representing people as a means of feedback (e.g. Benyon and Mival, 2008) omitted the impact of emotional expressions from these characters. The two studies in this paper attempted to address this gap and understand how to best use avatars to present positive and negative feedback as a motivational tool. This was achieved by, first, validating the visual avatars and the types of verbal feedback and, secondly, by showing when it is best to use visual or verbal and positive or negative feedback. The results suggest design implications by which avatars should express affect to users; positive feedback should be delivered via visual expressions for maximum impact. Negative feedback should be delivered via verbal feedback instead of visual expressions; this less evocative format may effectively convey meaning while avoiding highly suggestive, and potentially demotivating, content. We also suggest that in order to utilize avatars as a channel for motivational feedback, it is important to carefully select the captions and expressions that will be shown. Understanding the message the avatar is trying to convey is of paramount importance when designing feedback. Consistency between expression and message is also required in order for the user to easily interpret the feedback. This paper showed readily identifiable emotions and strong consistency between the captions and the expressions presented, which we argue allows the media presented to be used confidently in future motivational and persuasive avatar systems.

8.1. Limitations

One of the limitations of this work is the measures regarding persuasion and trustworthiness from the stimuli validation study. These questions were asked with a lack of context within the evaluation; therefore, participants may have had difficulty providing meaningful responses. Future work could address this by ensuring there is some contextual background or scenario presented to participants when asking them to

measure persuasion and trust in avatars. Another limitation is the inconsistency of the neutral expression with the different levels of messages in the comparison study (Noël *et al.*, 2009). This was an attempt to separate the expressions from the verbal messages, however, consistency is key to message understanding and this may have contributed to the higher happiness ratings for the consistent visual feedback. Future work could address this by comparing the text with more consistent expressions.

One other limitation of the work in this paper is the lab-based and short-term nature of the experimental studies. Fieldwork conducted over the long-term would add to the value of the experimental findings. An application could be developed for use in long-term behavior change scenarios that introduced gamification techniques via playful, expressive characters with the objective of capturing and maintaining users' long-term attention and engagement. Future work in this area should involve deploying and evaluating such a system in a long-term field study. This study could focus on changes in attitudes, behaviors and whether avatars can be utilized to increase motivation.

9. CONCLUSION

In sum, this paper contributes to the body of knowledge about how to design and use avatars as a means for providing feedback. It argues this type of presentation has the potential to engage users long-term, a key pressing issue for practical applications. Long-term engagement with users will be an important component of successful behavior change interventions. This paper has described two studies that validate a series of emotional avatar media and then contrast how it is interpreted from the perspective of motivation. Future work will focus on deploying a working system in an eco-feedback scenario and within people's homes in order to test the persuasiveness and trustworthiness of the feedback, as well as long-term engagement and behavior change. Such a persuasive system that encourages long-term engagement could have real effects on raising awareness and reducing negative behaviors.

ACKNOWLEDGEMENTS

The work described in this paper was conducted as part of the SINAIS project (CMU-PT/HuMach/0004/2008). This project was funded by the Portuguese National Science Foundation under the framework of the Carnegie Mellon Portugal Program. This was a 3-year research project led by the Madeira Interactive Technologies Institute at the University of Madeira in a consortium that involved Carnegie Mellon University, the Faculty of Engineering of the University of Porto, the Portuguese Catholic University and several companies and government agencies.

REFERENCES

- Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. (2005) A review of intervention studies aimed at household energy conservation. *J. Environ. Psychol.*, 25, 273–291.
- Bailenson, J.N., Blascovich, J., Beall, A.C. and Loomis, J.M. (2001) Equilibrium theory revisited: mutual gaze and personal space in virtual environments. *Presence*, 10, 583–598.
- Baker, S.C., Wentz, R.K. and Woods, M.M. (2009) Using virtual worlds in education: Second Life® as an Educational Tool. *Teach. Psychol.*, 36, 59–64.
- Baylor, A.L. (2011) The design of motivational agents and avatars. *Educ. Technol. Res. Dev.*, 59, 291–300.
- Baylor, A.L. and Kim, S. (2009) Designing nonverbal communication for pedagogical agents: when less is more. *Comput. Hum. Behav.*, 25, 450–457.
- Benyon, D. and Mival, O. (2008) Landscaping personification technologies: from interactions to relationships. In CHI'08 Extended Abstracts on Human Factors in Computing Systems (CHI EA'08). ACM, Florence, Italy, pp. 3657–3662.
- Blascovich, J. (2002) A theoretical model of social influence for increasing the utility of collaborative virtual environments. In Proceedings of the 4th International Conference on Collaborative Virtual Environments. ACM, Bonn, Germany, pp. 25–30
- Boberg, M., Piippo, P. and Ollila, E. (2008) Designing avatars. In Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts (DIMEA'08). ACM, Athens, Greece, pp. 232–239.
- Carroll, J.M. and Russell, J.A. (1996) Do facial expressions signal specific emotions? Judging emotion from the face in context. *J. Person. Soc. Psychol.*, 70, 205–218.
- Casanueva, J.S. and Blake, E.H. (2001) The effects of avatars on co-presence in a collaborative virtual environment. In Annual Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT2001). Pretoria, South Africa.
- Creed, C. and Beale, R. (2008) Psychological responses to simulated displays of mismatched emotion expressions. *Interact. Comput.*, 20, 225–239.
- Consolvo, S., Klasnja, P., McDonald, D.W., Avrahami, D., Froehlich, J., Legrand, L., Libby, R., Mosher, K., and Landay, J.A. (2008) Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. In Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08). ACM, New York, NY, USA, pp. 54–63.
- Curry, S.J., Wagner, E.H. and Grothaus, L.C. (1991) Evaluation of intrinsic and extrinsic motivation interventions with a self-help smoking cessation program. *J. Consult. Clin. Psychol.*, 59, 318–324.
- De Gelder, B. and Vroomen, J. (2000) The perception of emotions by ear and by eye. *Cogn. Emotion*, 14, 289–311.
- Deci, E.L. (1972) The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organ. Behav. Hum. Perform.*, 8, 217–229.
- Deci, E.L., Eghrari, H., Patrick, B.C. and Leone, D.R. (1994) Facilitating internalization: the self-determination theory perspective. *J. Pers.*, 62, 119–142.
- Deci, E.L., Koestner, R. and Ryan, R.M. (1999) A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol. Bull.*, 125, 627–668.
- Dillahunt, T., Becker, G., Mankoff, J. and Kraut, R. (2008) Motivating environmentally sustainable behavior changes with a virtual polar bear. Pervasive 2008 Workshop on Pervasive Persuasive Technology and Environmental Sustainability in Sydney, Australia.
- Dormann, C. (2000) Designing electronic shops, persuading consumers to buy. In Proceedings of the 26th Euromicro Conference. EUROMICRO 2000. Informatics: Inventing the Future, Vol. 2. IEEE Computer Society, pp. 140–147.
- Dyck, M., Winbeck, M., Leiberg, S., Chen, Y., Gur, R.C. and Mathiak, K. (2008) Recognition profile of emotions in natural and virtual faces. *PLoS One*, 3, e3628.
- Fabri, M., Grange, T., Park, B. and Dj, M. (2002) Expressive agents: non-verbal communication in collaborative virtual environments. AAMAS Workshop on Embodied Conversational Agents.
- Fogg, B.J. (2003). *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann, San Francisco.
- Froehlich, J., Dillahunt, T., Klasnja, P., Mankoff, J., Consolvo, S., Harrison, B. and Landay, J. (2009) UbiGreen: investigating a mobile tool for tracking and supporting green transportation habits. In Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI'09). ACM, Boston, MA, USA, pp. 1043–1052.
- Head, M., Hassanein, K. and Cho, E. (2003) Establishing eTrust Through Humanized Website Design. In 16th Bled eCommerce Conference eTransformation, pp. 365–382.
- Hong, P., Wen, Z. and Huang, T.S. (2002) Real-time speech-driven face animation with expressions using neural networks. *IEEE Transactions on Neural Networks*, 13, pp. 100–111.
- Kimura, H. and Nakajima, T. (2010) EcoIsland: A Persuasive Application to Motivate Sustainable Behavior in Collectivist Cultures. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NordiCHI'10). ACM, Reykjavik, Iceland, pp. 703–706.
- Koger, S.M. and Scott, B.A. (2007) Psychology and environmental sustainability: a call for integration. *Teach. Psychol.*, 34, 10–18.
- Krämer, N.C., Simons, N. and Kopp, S. (2007) The effects of an embodied conversational agent's nonverbal behavior on user's evaluation and behavioral mimicry. *Intelligent Virtual Agents. Lecture Notes in Computer Science*, Vol. 4722/2007. Springer, Berlin, Heidelberg, pp. 238–251.
- Lee, M.K., Kiesler, S. and Forlizzi, J. (2011) Mining Behavioral Economics to Design Persuasive Technology for Healthy Choices. *CHI*, pp. 325–334.
- Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G. and Strub, H.B. (2006) Fish 'n' Steps: encouraging physical activity with an interactive computer game. In Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp'06). Springer, Berlin, Heidelberg, pp. 261–278.

- Locke, E. A. and Latham, G. P. (2002) Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. *Am. Psychol.*, 57, 705–717.
- Mahmud, A.A., Dadlani, P., Mubin, O., Shahid, S., Midden, C. and Moran, O. (2007) iParrot: towards designing a persuasive agent for energy conservation. In *Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE'07)*. Springer, Berlin, Heidelberg, pp. 64–67.
- Nass, C. and Moon, Y. (2000) Machines and mindlessness: social responses to computers. *J. Soc. Issues*, 56, 81–103.
- Nass, C., Steuer, J. and Tauber, E.R. (1994) Computers are Social Actors. In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'94)*. ACM, Boston, MA, USA, pp. 72–78.
- Noël, S., Dumoulin, S. and Lindgaard, G. (2009) Interpreting human and avatar facial expressions. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I (INTERACT '09)*. Springer, Berlin, Heidelberg, pp. 98–110.
- Theng, Y.L. and Aung, P. (2011) Investigating effects of avatars on primary school children's affective responses to learning. *J. Multimodal User Interfaces*, 5, 45–52.
- von der Pütten, A.M., Krämer, N.C., Gratch, J. and Kang, S.-H. (2010) "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Comput. Hum. Behav.*, 26, 1641–1650.
- West, R. (2005) Time for a change: putting the transtheoretical (stages of change) model to rest. *Addiction*, 100, 1036–1039.
- Xbox LIVE Avatars <http://www.xbox.com/en-US/live/avatars>.
- Zimmerman, G.L., Olsen, C.G. and Bosworth, M.F. (2000) A 'stages of change' approach to helping patients change behavior. *Am. Family Physician*, 61, 1409–1416.