Designing Interface Agents: Beyond Realism, Resolution, and the Uncanny Valley

Eva Krumhuber[1], Marc Hall[2], John Hodgson[2], Arvid Kappas[1]

[1] School of Humanities and Social Sciences,
Jacobs University Bremen,
Campus Ring 1, 28759 Bremen,
Germany
[2] Department of Computing, Engineering and Technology,
University of Sunderland
St Peter's Way, Sunderland, SR6 0DD,
United Kingdom
{e.krumhuber, a.kappas}@jacobs-university.de
{marc.hall, john.hodgson}@sunderland.ac.uk

Abstract. Previous attempts in designing interface agents have been concerned mainly with producing highly realistic-looking animations with emotions that are clearly recognizable. We argue that the choice of visual representation requires consideration of purpose-related psychological processes (i.e., theory of mind) in users. In an evaluation study, four synthetic characters ranging in appearance from non-human to very human (blob, cat, cartoon, human) were evaluated with respect to dispositional traits, mental states, as well as emotions. Results showed that the type of synthetic character strongly influenced what judgment was made. Whilst the blob and cat characters were well liked, attributions of intelligence, mind and complex emotions were found to be reserved more for the human-like counterparts. The findings suggest that independently of questions of realism and clarity of emotional signs, the design of interface agents should be based on attributions the type of character elicits and the function the character is to serve in a particular application.

Keywords: interface agent, visual appearance, emotion, uncanny valley, theory of mind

1 Introduction

The visual appearance of computer agents and avatars is a topic of particular interest in the fields of computer science and AI. For the user, the anthropomorphic embodiment of a program for interaction appears much more tangible than the “black box” or a computational device displaying printed text on a screen. As such, the personal nature of its appearance allows for it to be more approachable and life-like, thereby making it an immediate source of interaction. This is not just a question of liking; but it changes the social relationship between users and agents/bots. From previous research we know that people attribute personality traits and human characteristics to interface agents similarly as they might do to other people [1]. In this sense, users respond emotionally to it and treat it as a social agent [2]. Considering this personification process, attempts have been made to increase the humanness of agents and avatars by adding human-like attributes. The ultimate goal of such endeavors consists for many developers in the creation of synthetic digital humans with photorealistic faces that exhibit life-like behavior [3], [4]. However, for practical reasons approaches are limited with respect to the type of realism that can be achieved [5]. Specifically, the design of such embodiments is driven by system constraints regarding the spatial and temporal resolution devices presently afford, as well as conceptual considerations. For example, anthropomorphic representations with high fidelity may lead to alienation as a consequence of the "uncanny valley effect" [6], [7]. Applying Mori's hypotheses which stem from a context in robotics to virtual agents one could argue that, if computer agents resemble humans too closely without making people fully believe that they are real, feelings of unpleasantness and uncanniness are triggered. In consequence, specific interactions might fail and users might try to avoid the "creepy" agent. To circumvent such pitfalls researchers have consequently chosen to implement the visual metaphor for their agents as cartoon-like humanoid characters, or animals, or animated objects, such as robots.

With regard to what these agents show, there has been a particular interest in the emotional expression. Emotions reveal much of a character's personality and influence the type and quality of interaction. For example, when users see a smiling agent they expect to have more enjoyable interactions compared to a non-smiling one [8]. The criterion here is typically whether an expression is recognized; this means whether a particular label, such as "happiness" is attributed to an entity when the designer intended to communicate this state. This concerns mainly how perceivers decode facial emotions, but it is not directly based on information on how senders would have encoded the expression in real life [9]. Furthermore, to maximize recognition, expressions are often not designed with ecological validity in mind. Thus, expressions correspond to stereotypical masks that are simplified in the type and quality of appearance. Mostly, these depict the six basic emotions (anger, fear, disgust, surprise, happiness, sadness) [10] and are displayed in a pure/exaggerated form [11]. Such expressions are well recognized because they function as clear representations of stereotypical emotion categories but they do not correspond to ecologically valid displays [12]. Furthermore, they do not necessarily capture the complexity of emotion attribution in the sense of what emotional states people really infer from the display [13].

Recent research in psychology may contribute in complicating the matter. Apparently, there is an interaction between how human we consider something and what mental and specifically emotional capacities we assume that "thing" to have [14], [15]. In other words, if something is less than human, we might not believe that it has the same mind a human has. Basic emotions, such as anger or happiness are easily attributed to animals, but more refined emotions, such as guilt or shame require more mind than we attribute to most animals [16]. Similarly, animated objects such as robots may remind of machines or automata and consequently lack emotions, cognitive flexibility and mind in the eye of the beholder. So what happens if agents range in appearance from highly anthropomorphic to cartoon-like or akin to animal, perhaps to escape the uncanny valley? Could it be that the type of representation affects perceivers in ways how (affectively) smart these beings are thought of? To elucidate such questions we conducted an evaluation study in which different types of visual agent representations – from non-human to very human - were presented. Depending on how closely the characters resemble humans, it was predicted that perceivers would make different attributions of dispositional traits, mental states and emotions. Furthermore, we investigated the effects of movement on characters' evaluation. Since Mori [6], [7] made different predictions concerning the slope of the uncanny valley for static and moving displays, attributions should change as a function of the display condition.

2 Evaluation Study

The study involved forty participants (21 men, 19 women) aged between 18-35 years (*M* = 20.33, *SD* = 2.96) who participated on a voluntary basis from Cardiff University, UK. All were students or staff at the university and received £7.00 for their participation. Participants were presented with either static or dynamic displays of four embodied characters that differed in their degree of humanness: blob, cat, cartoon, and human (see Fig. 1). In the static condition, images of the characters in a neutral position were shown for 5 s. In the dynamic condition, each character consecutively displayed three types of movement – idle, bow, wave – which lasted about 10 s. All characters were displayed on blue background with an image size of 490 x 270 pixels.

Fig. 1. Four embodied characters – blob, cat, cartoon, human - from non-human to very human in a neutral position.

Participants were tested individually on a PC workstation. After signing an informed consent form, they were told that they would see several animated characters that they should rate on a number of dimensions. It was made clear that there were no right or wrong answers. Rather, they should indicate their first impression. Using MediaLab 2008 (Empirisoft) software, participants could initiate the stimulus sequence by using the mouse to click a start button on the computer screen. Each stimulus randomly appeared for 5 s (images) or 10 s (videos) and was prefaced by a rating dimension that was displayed throughout the stimulus presentation. After the stimulus disappeared, participants were instructed to respond to the rating scale.

To allow for a varied nature in perception, we included a number of attributes that targeted dispositional traits, mental states as well as basic and social emotions. The following questions were answered on 7-point Likert-scales ranging from (1) *not at all* to (7) *very much*:

- How likeable is the character?
- How trustworthy is the character?
- How intelligent is the character?
- How engaging is the character?
- To what degree does the character have a mind on its own?
- To what degree can the character experience anger?
- To what degree can the character experience shame?

These questions were posed in random order, with one question per stimulus presentation.

3 Results

A multivariate analysis of variance (MANOVA) with condition (static, dynamic) and sex of participant (male, female) as between-subjects factors, and stimulus character (blob, cat, cartoon, human) as within-subjects factor was conducted on the seven dependent variables: likeable, trustworthy, intelligent, engaging, mind, anger, and shame. For all univariate analyses, a Greenhouse-Geisser adjustment to degrees of freedom was applied. There were no significant effects associated with sex of participant, $F(7, 30) = 0.98$, $p = .461$, and this factor was dropped in all further analyses. As expected, the multivariate main effect of stimulus character was highly significant, $F(7, 110) = 13.88$, $p = .000$. Univariate tests showed significant effects for nearly all variables: likeable, $F(2.56, 97.31) = 10.42$, $p = .000$, trustworthy, $F(2.87, 109.11) = 6.68$, $p = .000$, intelligent, $F(2.79, 106.21) = 14.97$, $p = .000$, engaging, $F(2.54, 96.41) = 1.80$, $p = .160$, mind, $F(2.75, 104.65) = 6.96$, $p = .000$, anger, $F(2.93, 111.32) = 15.98$, $p = .000$, and shame, $F(2.74, 104.32) = 5.26$, $p = .003$.

As can be seen in Fig. 2, for ratings of intelligence, the blob scored lowest and significantly different from the other characters (*p*s < .001). This was similar for attributions of mind in which the blob received lowest ratings which differed significantly from those of the cartoon ($p = .037$) and human character ($p = .005$). Furthermore, participants attributed less mind to the cat in comparison to the human character ($p = .003$). With respect to perceptions of anger, the cartoon character was judged to be most capable of experiencing anger with ratings significantly different from all other characters (*p*s < .01). Additionally, it was also perceived as least likeable and trustworthy, with ratings significantly lower than those of the remaining characters (*p*s < .05). The human character was perceived more capable to experience anger than the blob ($p = .05$). For ratings of shame, the human character scored significantly higher than both the blob ($p = .004$) and cat character ($p = .010$).

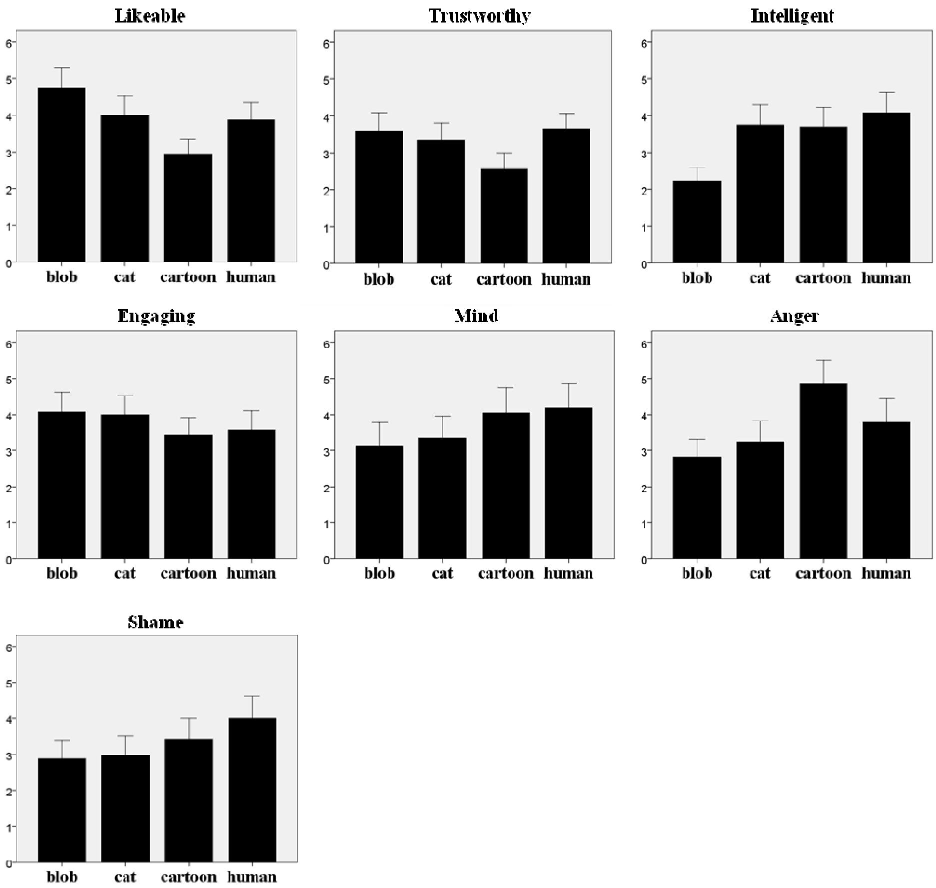

Fig. 2. Mean ratings of the four characters for the seven dependent measures. Error bars represent standard errors.

The multivariate main effect of condition was not significant, $F(7, 32) = 1.20$, $p = .331$. However, there was a significant interaction between condition and character, $F(7, 110) = 3.83$, $p = .001$. In univariate terms this interaction was significant only for ratings of likeability, $F(2.56, 97.31) = 4.57$, $p = .007$, and trustworthiness, $F(2.87, 109.11) = 3.03$, $p = .035$. Post-hoc comparisons showed that the cat character was perceived to be more likeable in the dynamic than in the static condition ($M_{\text{dynamic}} = 4.67$ vs. $M_{\text{static}} = 3.26$, $p = .006$). In contrast, trustworthiness ratings of the blob were significantly higher in the static than in the dynamic condition ($M_{\text{dynamic}} = 3.09$ vs. $M_{\text{static}} = 4.10$, $p = .042$).

4 Discussion

Results showed that the attribution of dispositional traits, mental states, as well as basic and social emotions differed depending on the type of computer agent. Overall, the blob was well liked, but ratings of intelligence and mind were lowest for this type of character. Given that it was the most non-human and object-like looking, participants might have ascribed less mental capacities which are usually reserved for humans [14], [15]. This is also reflected by the finding that the cat as a living, but non-human being was seen to possess less mind than the human character. Thus, the human appearance seems to play a crucial role in what kind of attributions people make. If something is less than human we might not perceive it as having the same mind as a human. Moreover, such lower perceived ability to reason and mentalize is interlinked with how emotionally smart those characters are seen. Specifically, refined emotions such as shame require more mind than what is attributed to objects and most animals. Respective ratings of the present study corroborate that notion. Both the blob and cat character were judged as being least capable to experience shame. In comparison, ratings of shame and anger were highest for the human and cartoon character, indicating that participants perceived them as being most capable to experience complex emotional and mental states.

For the proposed relation between human resemblance and perceiver's affinity, Mori [6], [7] had made slightly different predictions for moving and static displays. In the current study, attributions of likeability and trustworthiness were moderated by the type of display condition. Interestingly, this effect occurred for the two characters being furthest away from the human endpoint (i.e, blob and cat). Given that bowing and waving were chosen as representation of dynamic displays, it is feasible that these typical human movements exerted their influence particularly in how non-human characters were perceived. This is an intriguing finding as it suggests that the slope of the uncanny valley may not only be sensitive to the presence of motion, but also to the type of movement and how closely it represents human-like behavior.

5 Conclusion

The findings have important implications for the design of anthropomorphic characters in the field of computer science and AI. Previous efforts have focused largely on issues such as realism and emotional clarity. In that context, attempts have been made in producing highly realistic-looking animations with emotions that are easily recognizable [3], [11]. We argue that the design of agents is not just an issue of realism but requires consideration of purpose-related psychological processes in users. There is more to designing an agent than optimizing for the practical constraints of a particular implementation and avoiding the uncanny valley. It does make a difference whether an agent looks like a human, or an animal. It would appear that likeability is an important point, but if the blob is likeable but stupid, it would not be a good idea to use the blob to provide feedback in a serious matter. If the cartoon character is intelligent, but not trustworthy, you would not want to use such a representation in a sales-type interaction. In other words, depending on the function that a particular agent has, the choice of visual representation should take into account issues such as what types of inferences regarding the cognitive and emotional intelligence it invites. Here a closer collaboration of psychologists and computer scientists and engineers can be particularly promising. It would be interesting to what degree such effects persist over longer periods of interaction, or to what degree users of different ages (e.g, children) or from different cultural background are susceptible to such effects. More research is needed regarding these issues.

In psychology there is much research regarding Theory of Mind – this relates to the capacity of humans to imagine the thoughts and feelings of other humans [17]. When designing artificial interactants, whether embodied in the shape of robots, or virtual in the shape of agents, we must also consider the Theory of Mind the users are going to employ as a function of the design choices the engineers make [18], [19]. This study provides a pointer towards the type of evaluation studies that might be helpful in this context, but it is only a starting point towards the development of a systematic attempt to clarify criteria for development of artificial entities that can realize the communicative intent of its designers.

Acknowledgments. This work has been conducted within the European Commission project eCUTE – Education in Cultural Understanding, Technologically-Enhanced (FP7-ICT-2009.4.2). We thank Tony Manstead for his help with data collection.

References

1. Dehn, D.M., Van Mulken, S.: The Impact of Animated Interface Agents: A Review of Empirical Research. Int. J. Hum-Comput. St. 52, 1--22 (2000)
2. Reeves, B., Nass, C.: The Media Equation. Cambridge University Press, New York (1996)
3. Takács, B., Kiss, B.: The Virtual Human Interface: A Photorealistic Digital Human. IEEE Comput. Graph. 23, 38--45 (2003)
4. Alexander, O., Rogers, M., Lambeth, W., Jen-Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, Debevec, P.: The Digital Emily Project: Achieving a Photorealistic Digital Actor. IEEE Comput. Graph. 30, 20--31 (2010)
5. Donath, J.: Virtually Trustworthy. Science 317, 53--54 (2007)
6. Mori, M.: Bukimi No Tani. The Uncanny Valley (K. F. MacDorman & T. Minato, Trans.). Energy 7, 33--35 (1970)
7. Mori, M.: The Uncanny Valley. (K. F. Macdorman & N. Kageki, Trans.). IEEE Robot. Autom. Mag. 19, 98--100 (2012)
8. Sproull, L., Subramani, M., Kiesler, S., Walker, J.H., Waters, K.: When the Interface Is a Face. Hum-Comput. Interact. 11, 97--124 (1996)
9. Russell, J.A., Bachorowski, J.A., Fernández-Dols, J.M.: Facial and Vocal Expression of Emotion. Annu. Rev. Psychol. 54, 329--349 (2003)
10. Ekman, P.: All Emotions are Basic. In: Ekman, P., Davidson, R.J. (eds.): The Nature of Emotion: Fundamental Questions, pp. 15--19. Oxford University Press, New York (1994)
11. Kerlow, I.V.: The Art of 3D Computer Animation and Effects (3rd ed.). John Wiley & Sons, New Jersey (2004)

12. Kappas, A.: Smile When You Read This, Whether You Like It or Not: Conceptual Challenges to Affect Detection. IEEE T. Affect. Comput. 1, 38--41 (2010)
13. Krumhuber, E., Kappas, A.: Moving smiles: The Role of Dynamic Components for the Perception of the Genuineness of Smiles. J. Nonverbal. Behav. 29, 3--24 (2005)
14. Haslam, N., Loughnan, S., Kashima, Y., Bain, P.: Attributing and Denying Humanness to Others. Eur. Rev. Soc. Psychol. 19, 55--85 (2008)
15. Waytz, A., Epley, N., Cacioppo, J.T.: Social Cognition Unbound: Insights into Anthropomorphism and Dehumanization. Curr. Dir. Psychol. Sci. 19, 58--62 (2010)
16. Bilewicz, M., Imhoff, R., Drogosz, M.: The Humanity of What We Eat: Conceptions of Human Uniqueness Among Vegetarians and Omnivores. Eur. J. Soc. Psychol. 41, 201--209 (2011)
17. Frith U, Frith, C.D.: Development and Neurophysiology of Mentalizing. Philos. T. Roy. Soc. B. 358, 459--73 (2003)
18. von der Pütten, A., Krämer, N.: A Survey on Robot Appearances. In: 7th Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 267--268, ACM Press, New York (2012)
19. Hegel, F., Krach, S., Kircher, T., Wrede, B., Sagerer, G.: Theory of Mind (ToM) on Robots: A Functional Neuroimaging Study. In: 3rd ACM/IEEE International Conference on Human-Robot Interaction, pp. 335--342, ACM Press, New York (2008)