The Effect of Multimodal Emotional Expression and Agent Appearance on Trust in Human-Agent Interaction

Ilaria Torre
torrei@tcd.ie
ADAPT Research Centre
Trinity College Dublin
Dublin, Ireland

Emma Carrigan
carrige@tcd.ie
Trinity College Dublin
Dublin, Ireland

Rachel McDonnell
ramcdonn@tcd.ie
ADAPT Research Centre
Trinity College Dublin
Dublin, Ireland

Katarina Domijan
katarina.domijan@mu.ie
Maynooth University
Maynooth, Co. Kildare, Ireland

Killian McCabe
killianpmccabe@gmail.com
ADAPT Research Centre
Dublin, Ireland

Naomi Harte
nharte@tcd.ie
ADAPT Research Centre
Trinity College Dublin
Dublin, Ireland

ABSTRACT

Emotional expressivity can boost trust in human-human and human-machine interaction. As a multimodal phenomenon, previous research argued that a mismatch in the expressive channels provides evidence of joint audio-video emotional processing. However, while previous work studied this from the point of view of emotion recognition and processing, not much is known about what effect a multimodal agent would have on a human-agent interaction task. Also, agent appearance could influence this interaction too. Here we manipulated the agent's multimodal emotional expression ("smiling face" and "smiling voice", or both) and agent type (photorealistic or cartoon-like virtual human) and assessed people's trust toward this agent. We measured trust using a mixed-methods approach, combining behavioural data from a survival task, questionnaire ratings and qualitative comments. These methods gave different results: while people commented on the importance of emotional expressivity in the agent's voice, this factor had limited influence on trusting behaviours; while people rated the cartoon-like agent on several traits higher than the photorealistic one, the agent's style also was not the most influential feature on people's trusting behaviour. These results highlight the contribution of a mixed-methods approach in human-machine interaction, as both explicit and implicit perception and behaviour will contribute to the success of the interaction.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; • **Applied computing** → *Psychology*.

KEYWORDS

trust, virtual agents, multimodal emotional expression, mixed-methods

1 INTRODUCTION

Emotional expressivity – i.e., the ability to express and recognise emotions – is an essential aspect of social interaction [Russell et al. 2003], which helps inform others about a change in the environment and increases trust [Boone and Buck 2003; Oosterhof and Todorov 2009]. To make artificial agents more acceptable and integrated in future human-machine societies, it is therefore important that they are equally able to express and recognise human emotions.

Emotional expressivity is also multimodal by nature, as we make use of our face [e.g. Elfenbein and Ambady 2002], voice [e.g. Scherer et al. 2001], and bodily gestures [e.g. Meeren et al. 2005] to convey our emotional state. This multimodality has led researchers to study the presence of a multi-channel integration in emotion perception, e.g. by mismatching the emotion expressed in two channels. For example, [Mower et al. 2008] mismatched an animated character's face and an actress's voice expressing an emotion over semantically neutral utterances, which participants rated in terms of valence, activation and dominance. They found that the classification rate was highest for audio-visual and audio only stimuli, and lowest for video only. In a set of earlier experiments, [De Gelder and Vroomen 2000] similarly found that audio and video features influenced the perception of an audio-visual emotional expression, even when participants were specifically told to ignore one of the two channels. These studies hint at the possibility of audio-visual integration, in a sort of emotional McGurk effect [McGurk and MacDonald 1976] (see also [Creed and Beale 2008; Fagel 2006]). It has also been suggested that the audio and video channels express different components of emotions; specifically, the video channel mostly expresses the emotional valence (positive – negative) and the audio channel the

Figure 1: The (left) cartoon and (right) photorealistic agents with a smiling facial expression.

emotional activation (high excitation – low excitation) [Fagel 2006; Hanjalic 2006; Mower et al. 2009].

In another study by [Khooshabeh et al. 2013], people played a negotiation game where they had to sell phone plans to an avatar, which had a mismatched facial expression (happy/angry) and negotiation strategy (tough/soft). They found that people felt more threat (as indexed by lower cardiac output) when the avatar had angry expression but soft strategy, and similarly people looked more at the avatar's face in this condition. It might be because this combination sparks a 'reaction to the unknown', so people try to look at the primary source of emotional expression (the face) to make sense of what is happening and get more information [cf. Van Kleef et al. 2004].

These studies focused on the effect of this mismatch on people's ability to recognise and categorise the expressed emotions, whether as discrete categories [e.g. De Gelder and Vroomen 2000] or through multidimensional continua [e.g. Mower et al. 2009], but did not look at what effect this mismatch has on people's behaviour or overall perception of the emotionally expressive agent. While it is well established in Human-Machine Interaction that expressive agents are generally perceived as more trustworthy, engaging, and likeable [Elkins and Derrick 2013; Leite et al. 2012; Mathur and Reichling 2016; Tielman et al. 2014], it remains to be seen how a mismatch in an artificial agent's face and voice affects the human interactants. If, as mentioned in [Mower et al. 2009], people can consciously use a mismatch between their expressed and intended emotion in order to e.g. make a sarcastic remark, it is likely that users would not appreciate interacting with an artificial agent that presents a similar mismatch. Also, a limitation of previous studies [Mower et al. 2008, 2009] was that their artificial agent's face was limited in terms of expressivity.

Recently, [Torre et al. 2018a] looked at the mismatch of emotional expressivity of a photorealistic virtual human, and found that people did not overall trust or like this character very much. This might be because people find stylised or cartoon-like virtual humans more trustworthy and appealing than realistic ones. For example, a cartoon-like agent was rated as higher than a realistic one on Agreeableness – one of the Big Five traits, which includes trustworthiness [McCrae 2009] – in [Zibrek and McDonnell 2014]. In [McDonnell et al. 2012], people rated as high in appeal and trustworthiness characters that were either very cartoon-like or very realistic, but not characters on the middle ground. These studies

suggest that the style used to render a character can alter the interpretation of its personality, and that some realistic virtual agents might be located in the 'Uncanny Valley' [Mori 1970], creating a feeling of unease. In fact, in [Tinwell et al. 2013], uncanny virtual characters were also attributed more negative personality traits – including untrustworthiness. Similarly, in Human-Robot Interaction, previous studies found that people did not prefer or were engaged by more anthropomorphic robotic systems [Hastie et al. 2017; Lemaignan et al. 2015].

Therefore, we decided to run a similar experiment to [Torre et al. 2018a], adding a new condition with a second, cartoon-like virtual human. For these virtual humans, we chose two rendering styles that are typical of two types of entertainment industries: video games (realistic agent) and animated movies (cartoon-like agent). The two virtual humans were created using motion capture techniques, so that the face and voice of the resulting characters were equally expressive. In our experiment, people interacted with the cartoon and photorealistic virtual agents during a 'survival task' (Section 2.1). We manipulated the agent's emotional expressivity in terms of smiling: the agent was either *congruently* smiling in the voice and face, *congruently* neutral in the voice and face, or *incongruently* smiling only in the voice or only in the face. We focused on smiling as it is a universal [Mehu and Dunbar 2008] and multimodal [Eisenbarth and Alpers 2011] emotional expression, which is also audible from the voice alone [El Haddad et al. 2017], and which has been shown to increase credibility and trust [Elkins and Derrick 2013; Reed et al. 2018].

With these manipulations, we hope to shed light on the effect of a mismatched emotional expression of different virtual agents on people's behaviour and explicit perception.

2 METHOD

2.1 The survival task

We used the 'survival task' as an implicit measure of trust [Hall and Watson 1970]. Here we modified it to assess whether participants would accept the agent's suggestions as to which objects would be most important for survival in a hypothetical crash scenario. At the beginning of the game, participants are told that they have crashed either in the middle of the desert, or on the surface of the moon with only 6 intact objects, and that their base camp/mother ship is located 200 miles away. They have to rank the 6 intact objects in order of importance for survival. After they make the initial ranking, a virtual agent, which was 'originally meant to be their navigation assistant', appears and suggests a different ranking order (Table 1). For example, the agent places whichever object participants had placed at position 1 (most important), to position 5 (second least important); and so on. After the agent made these suggestions, participants could make their final decision, by providing a final ranking order. The comparison between the agent's ranking and the participants' final ranking was the basis for our analysis of behavioural trust.

The 6 objects for each of the two tasks (desert and moon) were chosen from bigger lists published by the National Park Foundation's Electronic Field Trip [1] and by NASA [Hall and Watson 1970],

---

[1] See e.g. https://studylib.net/doc/9847005/the-desert-survival-problem-expert-ranking-and-rationale

Table 1: Diagram showing the changes made by the virtual character to the participants' rankings of objects A-E.

Rank	Participant's ranking		Agent's ranking
1	[A] $\rightarrow$	positive sentence on [E]	$\rightarrow$[E]
2	[B] $\rightarrow$	positive sentence on [D]	$\rightarrow$[D]
3	[C] $\rightarrow$	neutral sentence on [C]	$\rightarrow$[C]
4	[D] $\rightarrow$	neutral sentence on [B]	$\rightarrow$[B]
5	[E] $\rightarrow$	negative sentence on [A]	$\rightarrow$[A]
6	[F] $\rightarrow$	negative sentence on [F]	$\rightarrow$[F]

respectively. In order to avoid objects which would obviously be very important (such as oxygen tanks on the moon) or very useless (such as an alcoholic beverage in the desert), we eliminated the top 2 and bottom 2 objects, and chose 6 objects between the remaining. These objects were: pocket knife, air map, sunglasses, book on edible animals, pistol, and flashlight for the desert task, and nylon rope, parachute silk, portable heating unit, milk, life raft, and receiver transmitter for the moon task.

2.2 Stimuli

We prepared a script encompassing all the utterances that the agent could speak during the survival task. As the agent's ranking depends on the participant's initial ranking, we prepared three sentences for each object, one describing the object in a positive way, one in a neutral way, and one in a negative way, for a total of 18 sentences (see Table 1). Thus, the agent would describe the object positively when it moved the object to positions 1 or 2, neutrally when it moved the object to positions 3 or 4, and negatively when it moved the object to positions 5 or 6 . The valence of the sentences as positive, neutral and negative was validated in a pre-test.

We then hired a male Irish actor to be recorded in our motion capture studio. The actor read the 18 pre-scripted sentences while keeping a neutral and smiling expression. The audio was recorded using a wireless microphone attached to his face. Several takes were recorded to ensure consistency. We used a 23-camera Vicon Vantage optical motion capture system for body motion capture and a Technoprops video-based head-mounted facial capture system to record the actor's facial expressions.

The virtual agents were created using state-of-the-art computer graphics technology for modelling, animating, and rendering. The photorealistic agent, comprising over 250 scans of a real actor's facial expressions, was created by the company 3Lateral. These scans were then carefully combined into a controllable facial rig, which could then be driven by the motion capture. The cartoon-like agent was a free high-end artist rig created by Artella. The actor's facial movements were then retargeted onto the models, using Faceware Tech software for the facial movement and inverse kinematics for the movement of the head. Finally, advanced shaders (e.g., subsurface scattering for the skin) were used to create the highly realistic appearance in Autodesk Maya 2018 software. The final characters are shown in Figure 1.

The audio recordings from the actor were processed using the software Audacity. First, a noise removal filter was applied to the

recordings; then the full audio file was segmented so as to obtain one file per utterance; finally, the individual sound files were amplitude-normalized. These files were then lip-synced to the individual, corresponding video files using the software Lightworks. Neutral sound files were lip-synced to neutral and smiling video files, and smiling sound files were lip-synced to smiling and neutral video files, to obtain the 4 desired experimental conditions: smiling voice + smiling face (V_sF_s); smiling voice + neutral face (V_sF_n); neutral voice + neutral face (V_nF_n); neutral voice + smiling face (V_nF_s).

2.3 Procedure

The experiment had a 2 (agent type: realistic or cartoon) by 4 (emotional expression: V_sF_s, V_sF_n, V_nF_n, or V_nF_s) by 2 (task: desert or moon) within-subject design (Figure 2). As participants only played two tasks (one with each type of agent and task), the emotional expression condition was semi-counterbalanced within participants, following a 'fractional-factorial' approach [Collins et al. 2009]. The agent type condition was fully counterbalanced, while the task order was randomised.

Participants first saw a page with a description of the task; then they were taken to the first part of the game, where they saw a grid with six empty positions, and six icons representing the objects they had to rank. Hovering with the mouse over the icons would reveal the name of the objects, and these could be placed in the grid by dragging-and-dropping. After they made their first ranking, the virtual agent videos were played: the first video that was played was the one describing participant's object #1, which the virtual character moved to position #5; therefore, this was described in a negative way (Table 1); and so on. While the videos were playing, the corresponding objects appeared in a second grid, in the positions chosen by the agent. After all 6 videos were played, a third grid appeared, and the 6 objects were available to the participants again for ranking. After participants made their final ranking, they were taken to a new page asking them to rate the virtual agent on 8 traits using a Likert scale ranging from 1 (= not at all) to 5 (= very much). The traits were: realism, appeal, eeriness, trustworthiness, knowledge, attractiveness, happiness and intelligence. These traits were chosen because they had previously been used in research on the perception of virtual characters [Ho and MacDorman 2010; McDonnell et al. 2012; Zell et al. 2015]. Finally, participants could leave a comment on the experiment in a free-text box. The total duration was approximately 15 minutes.

Thus, we collected different measures of participants' behaviour and perception: an implicit/behavioural measure of trust (the survival task), an explicit/quantitative measure (the ratings after the task) and an explicit/qualitative measure (the free-text comments at the end of the experiment). While the survival task might not be as engaging and interactive as a game where people are free to explore a digital environment, it warrants higher experimental validity, in that it is a scenario that can be easily controlled and replicated.
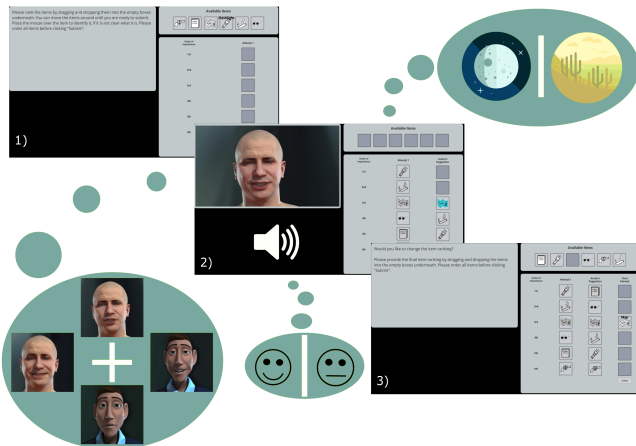
Figure 2: Diagram showing the experimental design: in panel 1) participants give their initial ranking, in the desert or moon scenario; in panel 2) the agent appears (in one of the agent type x emotional expression combinations) and suggests a different ranking; in panel 3) participants give their final ranking.

2.4 Participants

Participants were 84 students and staff from Trinity College Dublin, who were given a monetary compensation in exchange for participation. There were 58% males (N = 49), and the age range was 19-59 (median = 30, sd = 8.36); the majority of participants (N = 34) came from the Republic or Ireland, while the rest came from 26 other countries. Their self reported English fluency skills were as follows: 50% native (N = 42), 17% native-like (N = 14), 32% fluent (N = 27), 1% basic (N = 1).

3 RESULTS

3.1 Survival task results

First, we calculated the Kendal-Tau distance from the agent's ranking to the participants' final ranking (an integer going from 0 = smallest distance, to 11 = biggest distance). We fit a linear regression model with this distance as dependent variable, and virtual agent's emotional expression and type as predictors. This model was not significant, however. To unravel this model, we then looked at each object position separately: we calculated the change in rank for each of the objects. Since the agent always moves the objects to the same relative positions (Table 1), the change in ranks ranged from 0 to -5 for the object originally placed at position #1, from 1 to -4 for the object originally placed at position #2, from 3 to -2 for the object originally placed at position #4, and from 4 to -1 for the object originally placed at position #5. Then, we fitted a linear regression model on the change in ranks of each object, with agent emotional expression and agent type as predictors. Of these, only the change in ranks for the object originally placed at position #4 revealed a significant effect of emotional expression ($\chi^2(3) = 3.55, p = .016$); a post-hoc Tukey HSD test showed that the V_nF_n condition was trusted more than all the others (all $p \leq .05$). As shown in Fig. 3,
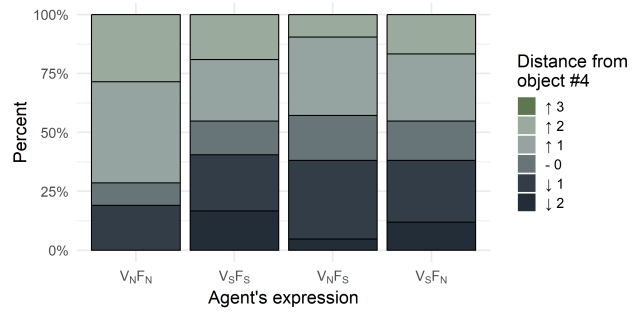

Figure 3: Changes in ranking from object #4 in the 4 emotional expression conditions.

more people repositioned their object #4 closer to the agent's position #2 in the V_nF_n condition than in any other (mean distance from V_nF_n = 1.81, from V_sF_s = 1.07, from V_nF_s = 1.10, from V_sF_n = 1.12).

Also, only the change in ranks for the object originally placed at position #2 revealed a significant effect of agent type ($\chi^2(1) = 5.64, p = .019$): more people repositioned their object #2 closer to the agent's position #4 when the agent was the cartoon-like one (mean distance from cartoon-like = -1.20, mean distance from photorealistic = -0.79)

3.2 Questionnaire results

We ran separate Kruskal-Wallis rank sum tests for each of the 8 traits (realism, appeal, eeriness, trustworthiness, knowledgeableness, attractiveness, happiness, intelligence), with emotional expression and agent type as predictors. Emotional expression was a significant factor only for the happiness ratings ($\chi^2(3) = 7.55, p = .05$); a post-hoc Dunn test of multiple comparisons, with Benjamini-Hochberg adjustment, showed that the V_sF_s condition was rated as happier than the V_nF_n ($Z = -2.64, p = .04$). On the other hand, agent type was a significant factor for realism ($\chi^2(1) = 40.34, p < .001$) and eeriness ($\chi^2(1) = 19.41, p < .001$), with the photorealistic agent scoring higher than the cartoon one; it was also significant for appeal ($\chi^2(1) = 12.97, p < .001$) and attractiveness ($\chi^2(1) = 20.19, p < .001$), with the cartoon agent scoring higher than the photorealistic one; and interestingly, the cartoon agent was also rated as happier than the photorealistic one ($\chi^2(1) = 8.56, p < .005$). The ratings which showed a significant difference based on agent type can be seen in Figure 4.

3.3 Qualitative results

Of the 84 participants, 22 left a free-text comment, either on one or on both games they played, for a total of 35 unique comments. Of these, 12 mentioned that they were not convinced by the agent's arguments, or that the information provided was insufficient to make a judgment. On the other hand, 8 people mentioned that the agent's suggestions made sense, and that their choices were affected by them.

Regarding emotional expressivity, two people commented that "The avatar was ... a little serious" and that "maybe should have changed from serious to slightly laughing when he made the joke

Figure 4: Average ratings of the traits that were rated significantly different based on agent type.

about the raft being useless. If a player had ranked the raft highly I think they would feel that the avatar was condescending". Both these people were referring to an agent in the V_sF_n condition. Another participant in the same condition said they were "undecided whether [the agent was] deceitful or not knowledgable". These comments suggest that the audio-visual mismatch might have conveyed the impression that the agent was not expressing a positive emotion (as the face was neutral), but rather a negative emotion with high activation, such as sarcasm [see Mower et al. 2009].

A few people remarked how the "avatar lacks the experience of ... how a human may think in such a situation", and that the "overall tone of the transmission was not really suited to an emergency situation", highlighting the importance of context. This has implications for the design of collaborative machines, as it shows that the appearance or voice of an agent should be designed also based on the task it will carry out [see also Torre et al. 2018b].

One participant remarked that "Regarding trustworthiness, I think the voice is a stronger factor than the avatar itself". This particular participant accepted all the suggestions from the agent in the first game played (with the photorealistic agent, V_nF_n condition, moon task), but not all suggestions in the second game (cartoon agent, V_nF_s condition, desert task). Another person said that "The voice of the avatar might also affect the perception. The avatar's voice did not sound like very knowledgeable". This participant had just played in the V_sF_n condition.

4 DISCUSSION

We combined several data collection methods (survival task, questionnaire, open-ended comments) to obtain an all-rounded view of participants' trust and explicit perceptions of a virtual agent, having manipulated the agent's emotional expression and appearance.

The two manipulations had a small influence on the ranking choices made in the survival task, suggesting that these features might not play the most important role when it comes to people's behaviour. Looking at people's comments, it emerged how other factors, such as the content of the agent's suggestions, might have influenced people's ranking decisions more. In [Torre et al. 2018a],

people tended to trust the agent in the V_nF_s condition more. The authors interpreted this as further evidence of joint audio-video emotional processing, as people might have trusted an agent expressing a positive valence (in the face) but not a high excitation (in the voice) [Fagel 2006; Hanjalic 2006; Mower et al. 2009], perhaps due to the disaster scenario they were put in. Here, we find some evidence that people trusted the V_nF_n agent more. These differences might be due to the setting in which the experiments took place: in the current study people played in a standard laboratory environment, which might have heightened a feeling of isolation and emergency, which led people to trust the 'serious' agent more. On the other hand, participants in [Torre et al. 2018a] were visitors of a museum who interacted with the experiment in the middle of an exhibition, and might have thus been less immersed in the danger of the simulated scenario.

While the agent's emotional expressivity did not have much impact on the survey responses, the agent type influenced all traits related to appearance (realism, attractiveness, appeal, eeriness), but not the traits related to performance / competence (trustworthiness, intelligence, knowledgeableness). From these ratings, it seems that people would explicitly prefer interacting with the cartoon-like agent. This is also evidenced by the effect found in the survival task, where people were more inclined to accept the suggestions from the cartoon-like agent in a certain part of the task. The fact that the cartoon-like agent was generally rated more positively than the realistic one supports previous findings [McDonnell et al. 2012; Tinwell et al. 2013; Zibrek and McDonnell 2014]. As the smiling and neutral utterances recorded from the actor varied naturally in their prosody and timing, it is possible that the manual lip-syncing produced some artefacts while mixing and matching the channels; perhaps, people would be more accepting of these in a cartoon-like virtual human, rather than a photorealistic one [McDonnell et al. 2012]. Also, a limitation of this study is that the models on which our two virtual characters were based are different, which could have caused a confound in the data (e.g. the photorealistic agent is bald, the cartoon agent has pointed chin, etc). However, here we were interested in testing the difference between two 'archetypical' virtual character styles. Previous findings on the difference between agent styles mainly reported implications for the computer graphics and entertainment industries [Geller 2008]; having added behavioural data to explicit ratings, we can add implications for the design of artificial agents in Human-Agent Interaction, as people seem to prefer stylised over photorealistic agent models. Furthermore, by using the 'survival task', a game that was originally devised to test leadership and negotiation skills [Hall and Watson 1970], our findings have specific applications for e.g. video games with a negotiation aspect, such as L.A. Noire. Mismatching a character's vocal and facial information, for example in terms of emotional expressivity, could be an interesting way to produce game characters that appear to be deceitful.

Interestingly, the agent type also influenced happiness, with participants rating the cartoon-like agent as happier than the photorealistic one. In [Zell et al. 2015], virtual characters with neutral or sad expressions were perceived as more realistic; here our cartoon character was perceived as happier than the photorealistic one; perhaps the realism goes both ways and a more realistic character is by default perceived to be more neutral in its expression, as cartoons

are generally meant to be exaggerating their expressions [Zell et al. 2015].

5 CONCLUSION

We found some evidence that people behaviourally trusted an agent with a congruent neutral expression. This might be because of the context of the interaction, which simulated an extreme survival scenario, and where expressivity could have been perceived as 'sarcastic' or 'inappropriate', as evidenced by participants' comments. Thus, the context of the Human-Agent Interaction might also play a role when designing interactive artificial agents, together with their appearance and voice. Also, people tended to behaviourally trust the cartoon-like agent more than the photorealistic one, and consistently rated it higher on several personality traits. This suggests that certain interactions might work better with stylised artificial agents rather than photorealistic ones, as these are more likely to cause an 'Uncanny Valley' effect. In all, we highlight the importance of collecting different kinds of data on the same phenomenon, as this allows to glimpse information that would otherwise be missed.

ACKNOWLEDGMENTS

REFERENCES

R. Thomas Boone and Ross Buck. 2003. Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior* 27, 3 (2003), 163–182.

Linda M. Collins, John J. Dziak, and Runze Li. 2009. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychological Methods* 14, 3 (2009), 202.

Chris Creed and Russell Beale. 2008. Psychological responses to simulated displays of mismatched emotional expressions. *Interacting with Computers* 20, 2 (2008), 225–239.

Beatrice De Gelder and Jean Vroomen. 2000. The perception of emotions by ear and by eye. *Cognition & Emotion* 14, 3 (2000), 289–311.

Hedwig Eisenbarth and Georg W. Alpers. 2011. Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion* 11, 4 (2011), 860.

Kevin El Haddad, Ilaria Torre, Emer Gilmartin, Hüseyin Çakmak, Stéphane Dupont, Thierry Dutoit, and Nick Campbell. 2017. Introducing AmuS: The amused speech database. In *International Conference on Statistical Language and Speech Processing*. Springer, 229–240.

Hillary Anger Elfenbein and Nalini Ambady. 2002. On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin* 128, 2 (2002), 203.

Aaron C. Elkins and Douglas C. Derrick. 2013. The sound of trust: Voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation* 22, 5 (2013), 897–913.

Sascha Fagel. 2006. Emotional McGurk Effect. In *Proceedings of the International Conference on Speech Prosody*, Vol. 1. ISCA.

Tom Geller. 2008. Overcoming the Uncanny Valley. *IEEE Computer Graphics and Applications* 28, 4 (2008), 11–17.

Jay Hall and Wilfred Harvey Watson. 1970. The effects of a normative intervention on group decision-making performance. *Human Relations* 23, 4 (1970), 299–317.

Alan Hanjalic. 2006. Extracting moods from pictures and sounds: Towards truly personalized TV. *IEEE Signal Processing Magazine* 23, 2 (2006), 90–100.

Helen Hastie, Katrin Lohan, Amol Deshmukh, Frank Broz, and Ruth Aylett. 2017. The Interaction Between Voice and Appearance in the Embodiment of a Robot Tutor. In *International Conference on Social Robotics*. Springer, 64–74.

Chin-Chang Ho and Karl F. MacDorman. 2010. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior* 26, 6 (2010), 1508–1518.

Peter Khooshabeh, Celso De Melo, Brooks Volkman, Jonathan Gratch, Jim Blascovich, and Peter Carnevale. 2013. Negotiation strategies with incongruent facial expressions of emotion cause cardiovascular threat. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 35.

Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2012. Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In *Proceedings of the 2012 Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 367–374.

Séverin Lemaignan, Julia Fink, Francesco Mondada, and Pierre Dillenbourg. 2015. You're doing it wrong! studying unexpected behaviors in child-robot interaction. In *International Conference on Social Robotics*. Springer, 390–400.

Maya B. Mathur and David B. Reichling. 2016. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition* 146 (2016), 22–32.

Robert R. McCrae. 2009. The five-factor model of personality. In *Handbook of Personality Psychology*, P. Corr and G. Matthews (Eds.). 148–161.

Rachel McDonnell, Martin Breidt, and Heinrich H. Bülthoff. 2012. Render me real?: Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 91.

Harry McGurk and John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264, 5588 (1976), 746.

Hanneke K. M. Meeren, Corné C. R. J. van Heijnsbergen, and Beatrice de Gelder. 2005. Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences* 102, 45 (2005), 16518–16523.

Marc Mehu and Robin I. M. Dunbar. 2008. Relationship between smiling and laughter in humans (Homo sapiens): Testing the power asymmetry hypothesis. *Folia Primatologica* 79, 5 (2008), 269–280.

M. Mori. 1970. Bukimi no tani [The uncanny valley]. *Energy*, 7(4), 33-35. (Translated by Karl F. MacDorman and Takashi Minato in 2005 within Appendix B for the paper Androids as an Experimental Apparatus: Why is there an uncanny and can we exploit it?. In *Proceedings of the CogSci-2005 Workshop: Toward Social Mechanisms of Android Science*. 106–118.

Emily Mower, Sungbok Lee, Maja J. Mataric, and Shrikanth Narayanan. 2008. Human perception of synthetic character emotions in the presence of conflicting and congruent vocal and facial expressions. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2201–2204.

Emily Mower, Maja J. Mataric, and Shrikanth Narayanan. 2009. Human perception of audio-visual synthetic character emotion expression in the presence of ambiguous and conflicting information. *IEEE Transactions on Multimedia* 11, 5 (2009), 843–855.

Nikolaas N. Oosterhof and Alexander Todorov. 2009. Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion* 9, 1 (2009), 128–133.

Lawrence Ian Reed, Rachel Stratton, and Jessica D. Rambeas. 2018. Face Value and Cheap Talk: How Smiles Can Increase or Decrease the Credibility of Our Words. *Evolutionary Psychology* 16, 4 (2018), 1–9.

James A. Russell, Jo-Anne Bachorowski, and José-Miguel Fernández-Dols. 2003. Facial and Vocal Expressions of Emotion. *Annual Review of Psychology* 54, 1 (2003), 329–349. https://doi.org/10.1146/annurev.psych.54.101601.145102

Klaus R. Scherer, Rainer Banse, and Harald G. Wallbott. 2001. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural Psychology* 32, 1 (2001), 76–92.

Myrthe Tielman, Mark Neerincx, John-Jules Meyer, and Rosemarijn Looije. 2014. Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 407–414.

Angela Tinwell, Deborah Abdel Nabi, and John P. Charlton. 2013. Perception of psychopathy and the Uncanny Valley in virtual characters. *Computers in Human Behavior* 29, 4 (2013), 1617–1625.

Ilaria Torre, Emma Carrigan, Killian McCabe, Rachel McDonnell, and Naomi Harte. 2018a. Survival at the Museum: A Cooperation Experiment with Emotionally Expressive Virtual Characters. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 423–427.

Ilaria Torre, Jeremy Goslin, Laurence White, and Debora Zanatto. 2018b. Trust in artificial voices: A "congruency effect" of first impressions and behavioural experience. In *Proceedings of APAScience '18: Technology, Mind,and Society (TechMindSociety '18)*.

Gerben A. Van Kleef, Carsten K. W. De Dreu, and Antony S. R. Manstead. 2004. The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology* 86, 1 (2004), 57.

Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To stylize or not to stylize?: The effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 184.

Katja Zibrek and Rachel McDonnell. 2014. Does render style affect perception of personality in virtual humans?. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, 111–115.