

Overcoming the Uncanny Valley: Displays of Emotions Reduce the Uncanniness of Humanlike Robots

Miriam Koschate*, Richard Potter*, Paul Bremner† and Mark Levine*

*Department of Psychology

University of Exeter, Exeter, EX4 4QG, U.K.

Email: m.koschate-reis@exeter.ac.uk, m.levine@exeter.ac.uk

†Bristol Robotics Lab

University of the West of England, Bristol, BS16 1QY, U.K.

Email: paul.bremner@uwe.ac.uk

Abstract—In this paper we show empirically that highly humanlike robots make thoughts of death more accessible, leading to perceptions of uncanniness and eeriness of such robots. Rather than reducing the humanlikeness of robots, our research suggests the addition of emotion displays to decrease a sense of uncanniness. We show that a highly humanlike robot displaying emotions in a social context reduces death-thought accessibility (DTA), which in turn reduces uncanniness. In a pre-test with $N = 95$ participants, we established that not all humanoid robots elicit thoughts of death and that the extent to which a robot appears humanlike may be linked to DTA. In our Main Study, $N = 44$ participants briefly interacted with a highly humanlike robotic head that either showed appropriate basic emotions or reacted by blinking. The display of emotions significantly reduced perceptions of uncanniness, which was mediated by a corresponding reduction in DTA. Implications for the design of humanoid robots are proposed.

Keywords—humanoid robot; uncanny valley; emotion display; death-thought accessibility (DTA)

I. INTRODUCTION

The *uncanny valley* has haunted humanoid robot design for several decades since Mori proposed the relationship between humanlikeness and familiarity in the 1970s (Fig. 1¹) [1]. With recent advances in anthromorphic design, humanlike robots are increasingly designed for the service industry and as social companions [2]. However, Mori suggested that while increasing humanlikeness increases familiarity, if a robot is too humanlike, it creates a sense of uncanniness reducing its familiarity [1]; an effect prevalent with the near perfect imitation of human appearance. This has led companies that develop and produce social robots to reduce human features to cartoon-like characteristics (e.g. Aldebaran’s *NAO* and *Pepper*) in order to hit the sweet spot of humanlikeness before ratings tumble into the *uncanny valley*.

To advance anthromorphic design, it is therefore crucial to understand the reasons underlying the *uncanny valley*. This enables us to propose solutions that are not merely based on

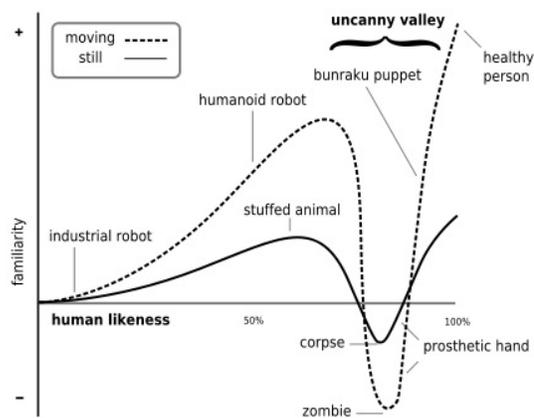


Fig. 1. The *uncanny valley* by Mori (1970)

assumptions [3]. In recent years, researchers have suggested a number of reasons why humanlike robots create a sense of eeriness and uncanniness. One hypothesis is that humanlike robots serve as reminders of death [4][5]. Empirical evidence shows that a picture of a highly humanlike robot elicited a stronger worldview defense - a typical reaction to death reminders - than a picture of a similar-looking human [4].

This association between humanlike robots and death may occur for several reasons: Firstly, humanlike robots are portrayed as threatening to humans in popular culture, both physically due to their super-human strength, as well as mentally [6]. Secondly, they may remind us of our creatureliness and associated mortality [7]. And third, their lack of emotions and empathy may remind us that they are, in fact, not alive despite looking alive. It is the third reason that we will focus on in this research.

Emotions have a strong social function [8]. They communicate fundamental ideas of cooperation, competition, conflict and social status. Displays of happiness, such as a smile, are universally understood as a sign of cooperation [9] whereas

¹Image used from https://commons.wikimedia.org/wiki/File:Mori_uncanny_valley.svg, under a creative commons licence

anger signals conflict [10]. Social emotions such as embarrassment, guilt and shame signal a shared understanding of norms and morality, which in turn engenders trust [11]. It is therefore unsurprising that emotions have been associated with being alive, and serve to differentiate humans from machines and objects [12]. Being able to express even basic emotions such as happiness, fear, and sadness, can signal human nature [12]. Human nature is defined as “[c]haracteristics that are typically or essentially human that represent the concept’s ‘core’” [12][p.256]. Concepts that are seen as representing human nature are emotional responsiveness, interpersonal warmth, cognitive openness, agency, and (emotional) depth. Accordingly, empirical evidence shows that humans and automata (e.g. robots) are judged differently on such traits [for a review of evidence see [13]].

Although the link between emotions and aliveness helps to explain the link between a lack of emotions and objectification, it does not imply uncanniness or death. Clearly, most machines and objects are not alive, but they are also not uncanny. For uncanniness to occur, a mismatch between expectation and perception is necessary [14]. In contrast to most objects and machines, humanlike robots create the expectation of human nature. This expectation is violated by a lack of emotional expression. They are ‘humans’ that are not alive. In a similar vein, zombies, clowns and masks are often perceived as uncanny as emotional expression is absent or masked. It is therefore not the knowledge of whether a robot can *detect* or *experience* emotions but rather its capability of *expressing* emotions that should reduce thoughts of death and hence uncanniness. In fact, knowing that a robot is capable of emotional experience but without being able to display emotions is likely to increase a sense of uncanniness as it creates a mismatch between expectation and perception.

II. RELATED WORK

In recent years, work on exploring the *uncanny valley* has intensified. Some researchers have focused on finding the *uncanny valley* by comparing robots of different degrees of humanlikeness [15] or by changing aspects of the face, such as the size and position of the eyes [16][17]. Although there are some indications that the *uncanny valley* exists, the shape of the curve at high levels of humanlikeness has been contested. For instance, it has been suggested that likeability may not recover to levels reached before the drop, and thus the curve may resemble more of an “uncanny cliff” [18]. However, findings from different studies cannot be easily compared as the outcome variable has been conceptualized in various different ways, ranging from familiarity and likeability, to eeriness and uncanniness. In our research, we will follow the suggestion by [19] and measure eeriness/uncanniness.

As mentioned earlier, recent work has explored the idea that uncanny robots remind us of death. Specifically, [4] compared reactions to an image of a humanlike robot and a similar-looking human. The study found that participants who had looked at the humanlike robot were more likely to show world-view defense by choosing a charismatic leader

based on a speech. World-view defense has been shown to be a reaction to death activation in prior research on terror management theory [20][21]. However, [4] found no direct effect of the pictures on death-thought accessibility (DTA). Detecting death-thought accessibility is highly dependent on the timing of the measurement. When death reminders are consciously processed, suppression sets in, and the effect can only be measured after a delay [21]. In contrast, when death reminders are subconsciously processed, effects need to be measured without delay as the effect dissipates once other stimuli are processed. In [4], a delayed measurement was given as it was unclear whether conscious or subconscious processing would be more likely. This may have, inadvertently, led to the effect not being found. However, the study still provides initial evidence that uncanny robots remind us of death. It remains, however, to be explored whether only uncanny/highly humanlike robots exert this effect, or whether *all* (humanoid) robots serve as reminders of death.

Research on emotions and the *uncanny valley* has found mixed results. One study [22] showed that a robot with an emotional gait was not perceived as uncanny by participants without autistic spectrum disorder (ASD), and only ASD participants reported reduced levels of familiarity. In contrast, a series of experiments [23] showed that a computer which is described as being able to feel emotions (and other ‘experiences’) is perceived as uncanny (Experiment 2). Similarly, a person described as being unable to experience emotions created a sense of uncanniness (Experiment 3). The difference in findings between [23] and [22] could be due to the fact that a mismatch in expectation and perception regarding emotions is created in the ‘emotional experience’ studies [23] but not in the ‘emotional gait’ study [22]. In the latter study, increased humanlikeness was matched by subtle indications of emotional expression via gait. This supports our proposition that humanlikeness raises expectations of human nature that need to be satisfied by emotional expression as a sign of human nature.

Initial indications that emotional expression and human nature are perceived to be linked when making judgments about robots stem from [24]. Participants who interacted with a *NAO* that expressed positive and negative emotions via sound and gestures rated its human nature to be higher than those who interacted with an unemotional *NAO*. In contrast, the intelligence of the robot’s response did not affect ratings of human nature.

III. RESEARCH QUESTION

In this research, we first examine the link between humanoid robots and death-thought accessibility. More specifically, we test whether some humanoid robots (e.g. those high in humanlikeness) elicit more DTA than other humanoid robots. Although [4] provides some indirect evidence for the association between highly humanlike robots and death, the study compares a picture of a humanlike robot and a human. We wish to extend this method by comparing several humanoid robots of differing characteristics, such as differing levels of

humanlikeness. This should give us a clearer insight into the link between the *uncanny valley* and death-thought accessibility (DTA). Once we find this link, we can test whether emotion displays in humanlike robots can reduce this effect, thereby reducing uncanniness. In contrast to [23], we do not focus on the (ascribed) experience of emotionality but instead on the communication aspect of emotions, that is facial emotion displays.

Hypothesis 1. Humanoid robots do not necessarily elicit DTA. Specifically, we predict that robots of high humanlikeness elicit more DTA than other humanoid robots.

Hypothesis 2. Emotion displays reduce the uncanniness of a highly humanoid robot compared with a neutral facial display.

Hypothesis 3. The effect of emotion displays on uncanniness are due to a reduction in DTA.

We will test Hypothesis 1 in a pre-test as this will allow us to move on to the question of how to reduce DTA (and uncanniness) in highly humanlike robots. Hypotheses 2 and 3 will be tested in the Main Study.

IV. CONTRIBUTION TO RESEARCH

Our paper aims to contribute to the literature in three important ways:

- 1) We seek to clarify whether DTA is only activated by highly humanlike robots, or whether DTA is an effect arising from perceiving a humanoid robot irrespective of other features of appearance (pre-test, Hypothesis 1). Previous research has established that a humanoid robot elicits more DTA than a human [4]. However, our study is the first to explore whether humanoid robots of differing appearance elicit DTA or mainly those considered highly humanlike.
- 2) This paper aims to provide initial evidence for the role of emotional displays in uncanniness. More specifically, we seek to show that emotional displays can effectively reduce DTA and uncanniness in highly humanlike robots (Main Study, Hypothesis 2). Such a discovery would be highly relevant for the design of humanlike robots.
- 3) Our research aims to establish for the first time whether DTA and perceptions of uncanniness are directly related (Main Study, Hypothesis 3). Establishing this link would contribute to identifying DTA as an underlying mechanism for the uncanny valley.

V. PRE-TEST: DEATH-THOUGHT ACCESSIBILITY IN DIFFERING HUMANOID ROBOTS

In a pre-test, we aimed to establish whether humanoid robots of differing characteristics elicit different levels of death-thought accessibility. In particular, we were interested in levels elicited by robots judged to be high in humanlikeness compared to those of lower humanlikeness.

A. Pre-ratings of material

To select the stimulus material for the pre-test, we first asked $N = 10$ participants to rate pictures of the faces of several humanoid robots (*Robovie-ii*, Honda's *ASIMO*, Aldebaran



Fig. 2. Set of humanoid robots used in the pre-test

Robotics' *NAO*, Aldebaran Robotics' *Pepper*, *KASPAR* [25], Hanson Robotics' *Jules*) with regard to their humanlikeness. The robot pictures were rated on a six-item semantic differential scale (e.g. human-made/human-like, artificial/natural) [19]. Answers were given on a 7-point scale. Based on the ratings of the pictures with regard to their humanlikeness, we selected three pictures for our pre-test (Fig.2).

- 1) **Low humanlikeness.** A photo of the head of Honda's *ASIMO* was shown. The head of *ASIMO* resembles a white round helmet and black visor. No individual facial features can be seen in the photo, which scored low on the humanlikeness scale in the pre-test ($M = 1.28, SD = 0.19$).
- 2) **Medium humanlikeness.** A photo of the head of Aldebaran's *Pepper* was shown. The head of *Pepper* is white and shows some cartoon-like facial features including large eyes, a small mouth, and a tiny nose. *Pepper* had a medium score on the humanlikeness scale in the pre-test ($M = 4.50, SD = 0.66$).
- 3) **High humanlikeness** A photo of Hanson Robotics' *Jules* was shown. *Jules* is a male robot head with features that resemble a human face to a high degree of accuracy, including eyebrows, a fake (Flubber) skin, and hair. In the pre-test, *Jules* scored highest with regard to humanlikeness ($M = 6.15, SD = 0.39$).

B. Method

An online experiment was conducted to examine the effect of different humanoid robots of varying degrees of humanlikeness on DTA. A total of $N = 95$ participants (57 women (60 %), 38 men; aged 18-56 years, $M = 28.07, SD = 8.95$), who were not previously exposed to robots through their studies or work (i.e. non-experts), took part in an online study. Participants were randomized to one of three conditions: low humanlikeness ($n = 32$), medium humanlikeness ($n = 31$), high humanlikeness ($n = 32$). After an information and consent page, participants were presented with the picture of the head of a humanoid robot for 30 secs (i.e. *ASIMO* in the low humanlikeness condition, *Pepper* in the medium humanlikeness condition, or *Jules* in the high humanlikeness condition). Participants were instructed to look at the picture and were told that they may be asked questions about the picture afterwards, in order to increase motivation. This was followed by a DTA assessment. Upon completion, participants were debriefed and given a chance to sign up for a raffle to win a £20 Amazon voucher.

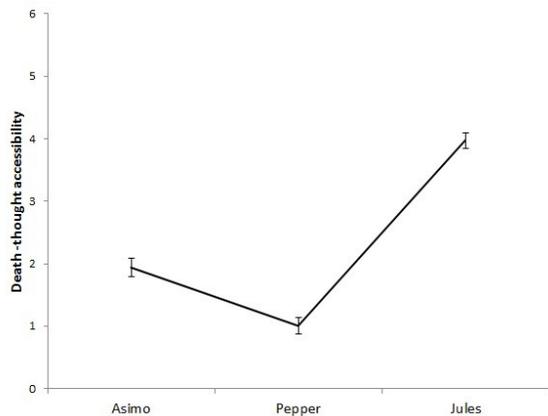


Fig. 3. Pre-test - Death-thought accessibility (0-6 death-related words) as a function of humanlikeness of robot pictures. Error bars represent 1 standard error (SE).

A word fragment task was used to measure DTA [21]. This measure is a standard measure of DTA, and has been widely used (see [26] for a review). The word fragment task infers highly active mental contents on the basis of responses to an ambiguous task. The idea is that active mental content influences the interpretation of ambiguous stimuli, thereby increasing the likelihood of the activated content becoming the dominant response [27]. Participants are given a list of 20 word fragments (e.g. C O F F _ _), of which six can be completed by either using words associated with death (buried, dead, grave, killed, skull, coffin) or words not associated with death (e.g. coffee instead of coffin). The number of death-related words that were provided by participants was summed to a DTA score, with higher scores representing higher DTA.

C. Results

A one-way ANOVA with DTA as the dependent variable revealed a significant effect of humanlikeness on DTA, with $F(2, 92) = 131.80, p < .001$. A post-hoc Tukey test showed that *Jules* ($M = 3.97, SD = 0.70$), the robot head with the highest level of humanlikeness, elicited significantly more DTA than both *ASIMO* ($M = 1.94, SD = 0.80, p < .001, Cohen's d = 2.71$) and *Pepper* ($M = 1.00, SD = 0.73, p < .001, d = 4.17$), respectively. *ASIMO*, the robot head with the lowest humanlikeness ratings, elicited significantly more DTA than *Pepper* ($p < .001, d = 1.23$), the robot head with medium levels of humanlikeness. The study therefore supports Hypotheses 1 that humanoid robots differ in the level of DTA they elicit. The pre-test also gives a first indication that high levels of humanlikeness create particularly high levels of DTA.

D. Conclusion

The pre-test provides first evidence that not all humanoid robots elicit DTA to the same extent. Indeed, as predicted, only a highly humanlike robot elicited substantial levels of DTA. Therefore, the pre-test provides some initial indication that DTA is related to high humanlikeness, and thus provides

some justification for the Main Study. However, the stimulus material used in the pre-test differs in more respects than just humanlikeness, and findings are therefore not conclusive. In the Main Study, we will use the same robot in each condition to make findings more comparable. Furthermore, the pre-test does not provide data to show a relationship between DTA and uncanniness. We will measure the link between DTA and ratings of uncanniness in the Main Study. Furthermore, the Main Study aims to test whether displays of emotion in a simple social interaction can reduce DTA and uncanniness in a highly humanlike robot.

VI. MAIN STUDY: EMOTION DISPLAY, DTA AND UNCANNINESS

In the Main Study, we examine whether displays of emotions in a social context can reduce uncanniness by reducing DTA. An experiment with one between-subjects factor (emotion display: yes/no) was conducted.

A. Method

A controlled experiment was conducted to examine the effect of emotion displays on uncanniness and DTA. A total of $N = 44$ participants (24 women (55%), 20 men; aged 22-58 years, $M = 32.84, SD = 9.62$) were recruited. Participants were randomized to either an emotion display condition ($n = 22$) or control condition ($n = 22$).

1) **Material:** A humanoid robot head, *Jules*, developed by Hanson Robotics, was used in the Main Study. *Jules* is capable of displaying a range of emotions. As can be seen from the pre-test data, *Jules* elicited high levels of DTA. This will allow us to test whether emotion displays can reduce both uncanniness and DTA in highly humanlike robots.

- 1) **Emotion display.** A positive emotion (happy) and a negative emotion (sad) were performed by the robotic head in response to a brief success or failure story by the participant. The emotion always matched the valence of the story, that is, a display of happiness followed a story of success, and a display of sadness was shown in response to a story of failure (Fig. 4).
- 2) **No emotion** The robot head only reacted by blinking to both the success and failure story. The face showed a neutral expression. This allowed us to control for movement without displaying an emotion.

As a manipulation check, participants were asked to rate to what extent the robotic head had expressed nine different emotions on a scale from 1 (*disagree*) to 5 (*agree*). Items 'happy' and 'sad' were target items for the manipulation check whereas the other seven items served as distractor items.

DTA was assessed in the same format as in the pre-test but as a paper-and-pencil version. Twenty word fragments were presented, out of which six word fragments served to measure DTA [21].

Uncanniness/eeriness was assessed with a semantic differential consisting of eight items (e.g. reassuring/eerie) [19], with a 7-point answering scale. Internal consistency of the scale was good with a Cronbach's α of .81.



Fig. 4. Jules performing a sad and happy display (Main Study)

Humanlikeness was assessed as a control variable. A six-item semantic differential scale (e.g. human-made/human-like, artificial/natural) was used [19]. Answers were given on a 7-point scale. The scale showed good reliability with a Cronbach's α of .85.

2) *Procedure*: Each participant was asked to sit opposite the robotic head which was placed on a table. Participants were given an information sheet and consent form. Once they had consented to participate, they were asked to think of a moment where someone they know had experienced a success/failure, and to briefly (in 2-3 sentences) tell the robotic head about the situation. In both conditions, each participant told both a success and failure story. The order of the story was counter-balanced across participants in both conditions so that half of participants told the success story first, and the other half told the failure story first. Once the story came to an end, the experimenter, who sat behind a screen and pretended to work on an unrelated task, triggered the response in Jules via the laptop keyboard in a 'Wizard-of-Oz' setup. Participants rated the emotion shown by Jules after each story. When the participant had finished both interactions, they were given a questionnaire containing the DTA word fragment task and items to measure uncanniness and humanlikeness. Participants were debriefed at the end of the study and were paid £2.50 for their participation.

B. Results

1) *Manipulation check*: First, we tested whether participants in the emotion display condition had recognized the intended emotion (happy/sad) correctly. We conducted a mixed-method ANOVA with condition (emotion display/control) as a between-subjects factor and story (success/failure) as a within-subjects factor. This analysis yielded a main effect of condition on happiness ratings, with $F(1, 42) = 23.73, p < .001$. It also showed a main effect of story, with $F(1, 42) = 228.62, p < .001$, and a condition by story interaction, with $F(1, 42) = 120.60, p < .001$. As expected, participants rated Jules to display happiness significantly more

in the emotion display condition after they had told a success story ($M = 4.82, SD = 0.39$) than after a failure story ($M = 1.09, SD = 0.29, p < .001, d = 10.85$), and significantly more than in the no emotion control condition, irrespective of story (control/success: $M = 2.14, SD = 1.46, p < .001, d = 2.51$; control/failure: $M = 1.55, SD = 0.91, p < .001, d = 4.67$).

A mixed-method ANOVA with sadness as dependent variable also showed the intended pattern, with a significant main effect of condition, with $F(1, 42) = 23.01, p < .001$, a significant main effect of story, with $F(1, 42) = 378.00, p < .001$, and a significant interaction, with $F(1, 42) = 298.67, p < .001$. Jules was rated as displaying significantly more sadness in the emotion display condition when a story of failure was told ($M = 4.91, SD = 0.29$) than after a success story ($M = 1.05, SD = 0.21, p < .001, d = 15.25$), and significantly more than in the control condition, regardless of story (control/success: $M = 1.73, SD = 1.20, p < .001, d = 3.64$; control/failure: $M = 1.95, SD = 1.17, p < .001, d = 3.47$). Thus, results show that participants had no problems identifying displays of happiness and sadness performed by the robotic head, Jules.

2) *Hypothesis 2*: Gender did not affect ratings of uncanniness ($t(42) = 0.66, p = .511$), humanlikeness ($t(42) = 1.09, p = .284$) or DTA ($t(42) = 0.93, p = .357$). However, age was related to lower ratings of uncanniness ($r(44) = -.46, p = .002$) but not to humanlikeness ($r(44) = -.10, p = .502$) or DTA ($r(44) = -.15, p = .321$). We will therefore control for age in subsequent analyses.

To test Hypothesis 2, that emotion displays reduce uncanniness, we performed an ANCOVA with condition (emotion display/control) as factor, uncanniness as dependent variable and age as a co-variate. Condition significantly affected uncanniness, with $F(1, 41) = 4.14, p = .048$. Participants in the emotion display condition ($M = 4.22, SD = 0.63$) reported significantly lower levels of uncanniness than participants in the control condition ($M = 4.73, SD = 1.02, d = 0.60$). This result supports our hypothesis that emotion displays reduce a sense of uncanniness/eeriness when interacting with a highly humanlike robot.

3) *Hypothesis 3*: To test Hypothesis 3, that emotion displays reduce uncanniness because they reduce DTA, we ran an ANCOVA with condition (emotion display/control) as factor, DTA as dependent variable and age as a co-variate. As expected, condition affected DTA significantly, with $F(1, 41) = 190.03, p < .001$. The emotion display condition ($M = 0.95, SD = 0.65$) elicited significantly lower levels of DTA than the control condition ($M = 4.27, SD = 0.94, d = 4.12$). This result shows that emotion displays, including a display of sadness, reduced DTA substantially. DTA is also significantly related to uncanniness, with $r(44) = .44, p = .003$.

In the next step, we tested whether the effect of condition on uncanniness is due to a reduction in DTA.

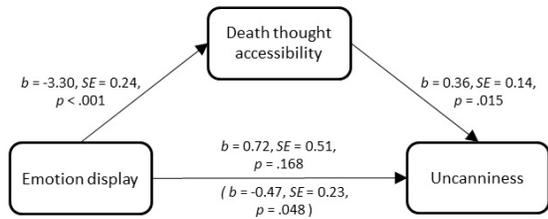


Fig. 5. Mediation effect (Main Study)

We used the PROCESS macro [28] to test for mediation using 5,000 bootstrap samples, controlling for age. Results support Hypothesis 3 (see Fig. 5). The indirect effect of condition on uncanniness via DTA was significant, with $b = -1.19$, $SE = 0.47$, $CI95 = [-2.05; -0.20]$, $Sobel z = -2.05$, $p = .012$. The significant effect of emotion display condition on uncanniness becomes non-significant once DTA is controlled for. The indirect path shows that displays of emotion reduce DTA significantly, and that DTA, in turn, affects uncanniness significantly.

In contrast to the effect on DTA, we found no effect of condition on ratings of humanlikeness ($t(42) = 0.60$, $p = .552$), emotion display: $M = 3.39$, $SD = 0.95$, control: $M = 3.58$, $SD = 1.21$), suggesting that emotion displays do not reduce uncanniness by increasing humanlikeness.

C. Conclusion

The Main Study provides support for the idea that thoughts of death and uncanniness are triggered by the mismatch between a humanlike appearance and a lack of human nature in the form of emotional expression. Once a highly humanlike robot displays emotions, DTA is reduced virtually completely, and uncanniness ratings are reduced significantly. In contrast, the addition of emotion displays did not affect perceptions of humanlikeness.

VII. DISCUSSION

The results of our experimental studies provide promising evidence that displays of emotions can reduce the uncanniness of highly humanoid robots. Furthermore, we provide initial evidence that uncanniness is, at least to some extent, due to an increase in death-thought accessibility.

Our results confirm our primary hypothesis that highly humanlike robots act as subtle reminders of death, creating a sense of eeriness and uncanniness. This replicates and extends findings reported by [4]. We provide initial, though not conclusive, evidence that highly humanlike robots trigger thoughts of death more than less humanlike robots, and that not all humanoid robots trigger thoughts of death to the same extent. More importantly, our data provide first evidence that a lack of emotion displays is the reason why humanlike robots remind us of death and appear uncanny.

However, it is important to bear in mind that these findings are initial evidence rather than conclusive results. Our

results on emotion displays were obtained with a single robot platform and a single experimental paradigm. Furthermore, while emotion displays reduced uncanniness significantly, it did not eliminate it. In contrast, DTA was virtually eliminated. This finding suggests that DTA is only a partial explanation for uncanniness, and that other factors are likely to play a role. This pattern of results also fits with the proposal that the *uncanny valley* is shaped more like a cliff with ratings not fully recovering once the low point has been reached [18].

VIII. DESIGN IMPLICATIONS

Though more research is clearly required to replicate and refine our findings, they suggest that anthropomorphic design should consider emotion displays in order to close the gap between the expectations the design raises with regard to human nature and the perception of it. Although facial expression of emotions are the easiest to recognize [29], other ways of expression may serve the same purpose. For instance, emotional gait [22], voice [30], or gestures, may be able to reduce DTA and uncanniness.

IX. LIMITATIONS AND FUTURE WORK

Although our experimental paradigm was successful in showing that emotion displays reduce DTA and uncanniness, it was also limited.

The first issue is that our design does not allow us to differentiate between the effect of the positive emotion display (happiness) and negative emotion display (sadness). However, even with the limited data we currently have, it is encouraging to see that sadness, an emotion that is closely associated with death, can reduce death-thought accessibility, or at least not undermine the effect of a positive emotion.

A second issue is that it remains unclear whether a social interaction - even a very brief one as in this study - is necessary for the effect to occur or whether simply perceiving the displayed emotion without social context would suffice. More specifically, the matching of the valence of the emotion to the valence of the story told by the participant creates a situation where the robot appears to show empathy. Empathy is perceived as a core element of humanness [12], and may thus be the key aspect that helped to reduce the mismatch between expectation and perception of human nature in our study.

ACKNOWLEDGMENT

This research is funded by the EPSRC under its IDEAS Factory Sandpits call on Digital Personhood, grant ref: EP/L00416X/1.

REFERENCES

- [1] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley," *Robotics & Automation Magazine, IEEE*, vol. 19, no. 2, pp. 98–100, 2012.
- [2] R. Yamazaki, S. Nishio, K. Ogawa, and H. Ishiguro, "Teleoperated android as an embodied communication medium: A case study with demented elderlies in a care facility," in *RO-MAN*. IEEE, 2012, pp. 1066–1071.
- [3] F. E. Pollick, "In search of the uncanny valley," in *User Centric Media*. Springer, 2010, pp. 69–78.

- [4] K. F. MacDorman, "Mortality salience and the uncanny valley," in *5th IEEE-RAS*. IEEE, 2005, pp. 399–405.
- [5] K. F. MacDorman and H. Ishiguro, "The uncanny advantage of using androids in cognitive and social science research," *Interaction Studies*, vol. 7, no. 3, pp. 297–337, 2006.
- [6] M.-S. Kim and E.-J. Kim, "Humanoid robots as the cultural other: Are we able to love our creations?" *AI & Society*, vol. 28, no. 3, pp. 309–318, 2013.
- [7] J. L. Goldenberg, "The body stripped down: An existential account of the threat posed by the physical body," *Current Directions in Psychological Science*, vol. 14, no. 4, pp. 224–228, 2005.
- [8] B. Parkinson, "Emotions are social," *British Journal of Psychology*, vol. 87, no. 4, pp. 663–684, 1996.
- [9] J. P. Scharlemann, C. C. Eckel, A. Kacelnik, and R. K. Wilson, "The value of a smile: Game theory with a human face," *Journal of Economic Psychology*, vol. 22, no. 5, pp. 617–640, 2001.
- [10] E. A. Van Doorn, M. W. Heerdink, and G. A. Van Kleef, "Emotion and the construal of social situations: Inferences of cooperation versus competition from expressions of anger, happiness, and disappointment," *Cognition & Emotion*, vol. 26, no. 3, pp. 442–461, 2012.
- [11] J. L. Tracy, R. W. Robins, and J. P. Tangney, *The self-conscious emotions: Theory and research*. Guilford Press, 2007.
- [12] N. Haslam, "Dehumanization: An integrative review," *Personality and Social Psychology Review*, vol. 10, no. 3, pp. 252–264, 2006.
- [13] N. Haslam, S. Loughnan, Y. Kashima, and P. Bain, "Attributing and denying humanness to others," *European Review of Social Psychology*, vol. 19, no. 1, pp. 55–85, 2008.
- [14] E. Jentsch, "On the psychology of the uncanny (1906)," *Angelaki: Journal of the Theoretical Humanities*, vol. 2, no. 1, pp. 7–16, 1997.
- [15] K. F. MacDorman, "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley," in *ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*, 2006, pp. 26–29.
- [16] K. F. MacDorman, R. D. Green, C.-C. Ho, and C. T. Koch, "Too real for comfort? uncanny responses to computer generated faces," *Computers in Human Behavior*, vol. 25, no. 3, pp. 695–710, 2009.
- [17] J. Seyama and R. S. Nagayama, "The uncanny valley: Effect of realism on the impression of artificial human faces," *Presence: Teleoperators and Virtual Environments*, vol. 16, no. 4, pp. 337–351, 2007.
- [18] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita, "Is the uncanny valley an uncanny cliff?" in *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*. IEEE, 2007, pp. 368–373.
- [19] C.-C. Ho and K. F. MacDorman, "Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1508–1518, 2010.
- [20] F. Cohen, S. Solomon, M. Maxfield, T. Pyszczynski, and J. Greenberg, "Fatal attraction the effects of mortality salience on evaluations of charismatic, task-oriented, and relationship-oriented leaders," *Psychological Science*, vol. 15, no. 12, pp. 846–851, 2004.
- [21] J. Greenberg, T. Pyszczynski, S. Solomon, L. Simon, and M. Breus, "Role of consciousness and accessibility of death-related thoughts in mortality salience effects," *Journal of Personality and Social Psychology*, vol. 67, no. 4, p. 627, 1994.
- [22] M. Destephe, M. Zecca, K. Hashimoto, and A. Takanishi, "Uncanny valley, robot and autism: Perception of the uncanniness in an emotional gait," in *Robotics and Biomimetics (ROBIO), 2014 IEEE International Conference on*, Dec 2014, pp. 1152–1157.
- [23] K. Gray and D. M. Wegner, "Feeling robots and human zombies: Mind perception and the uncanny valley," *Cognition*, vol. 125, no. 1, pp. 125–130, 2012.
- [24] J. Złotowski, E. Strasser, and C. Bartneck, "Dimensions of anthropomorphism: From humanness to humanlikeness," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 2014, pp. 66–73.
- [25] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A. Mirza, and M. Blow, "Kaspar—a minimally expressive humanoid robot for human–robot interaction research," *Applied Bionics and Biomechanics*, vol. 6, no. 3–4, pp. 369–397, 2009.
- [26] J. Hayes, J. Schimel, J. Arndt, and E. H. Faucher, "A theoretical and empirical review of the death-thought accessibility concept in terror management research," *Psychological Bulletin*, vol. 136, no. 5, pp. 699–739, 2010.
- [27] E. Tulving, D. L. Schacter, and H. A. Stark, "Priming effects in word-fragment completion are independent of recognition memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 8, no. 4, pp. 336–342, 1982.
- [28] A. F. Hayes, *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press, 2013.
- [29] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 2004, pp. 205–211.
- [30] C. Breazeal, "Emotive qualities in robot speech," in *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, vol. 3, 2001, pp. 1388–1394.