Too Much Humanness for Human-Robot Interaction: Exposure to Highly Humanlike Robots Elicits Aversive Responding in Observers

Megan Strait[1,2], **Lara Vujovic**[2], **Victoria Floerke**[2], **Matthias Scheutz**[1], **and Heather Urry**[2]
Departments of [1]Computer Science and [2]Psychology
Tufts University, Medford MA 02155 USA
{megan.strait, lara.vujovic, victoria.floerke, matthias.scheutz, heather.urry}@tufts.edu

ABSTRACT

People tend to anthropomorphize agents that look and/or act human, and further, they tend to evaluate such agents more positively. This, in turn, has motivated the development of robotic agents that are humanlike in appearance and/or behavior. Yet, some agents – often those with highly humanlike appearances – have been found to elicit the opposite, wherein they are evaluated more negatively than their less humanlike counterparts. These trends are captured by Masahiro Mori's *uncanny valley hypothesis*, which describes a (uncanny) valley in emotional responding – a switch from affinity to dislike – elicited by agents that are "too humanlike".

However, while the valley phenomenon has been repeatedly observed via subjective measures, it remains unknown as to whether such evaluations reflect a potential impact to a person's behavior (i.e., aversion). We attempt to address this gap in the literature via a novel experimental paradigm employing both traditional subjective ratings, as well as measures of peoples' behavioral and phsyiological responding. The results show that not only do people rate highly humanlike robots as uncanny, but moreover, they exhibit greater avoidance of such encounters than encounters with less humanlike and human agents. Thus, the findings not only support Mori's hypothesis, but further, they indicate the valley should be taken as a serious consideration for peoples' interactions with humanlike agents.

Author Keywords

Human-robot interaction; uncanny valley; emotion regulation; situation selection/modification; attentional deployment; anthropomorphism; embodied conversational agents; virtual agents

ACM Classification Keywords

H.4 Information Systems Applications: Miscellaneous

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea

http://dx.doi.org/10.1145/2702123.2702415

INTRODUCTION

Based on the "like me" premise that people respond more positively towards agents similar to themselves ([27]), development of humanlike agents has been of growing interest to researchers in human-robot and human-computer interaction alike. For both robots and virtual agents, humanlike appearances and/or behaviors have been found to improve perceptions and increase rapport (e.g., [1, 9, 40]). Thus, by capitalizing on traits that are more familiar and intuitive to people, humanlike agents can offer more natural and effective interactions than their less humanlike counterparts [11].

However, agents that are "too humanlike" can produce the opposite effect. For instance, a recent exhibit by artist Tony Matelli – the Sleepwalker – was a highly realistic statue intended to depict a man "who is hopelessly lost" and "vulnerable"[1] amidst the open landscape of Wellesley College (see Figure 1). While the statue itself is obviously inanimate and therefore unable to do any physical harm, the unnerving nature of the statue's appearance resulted in a petition[2] that garnered over 1000 signatures for its removal – so that people "don't have to *confront him* as they go about their daily lives".

This response to the statue is consistent with Mori's uncanny valley hypothesis (UVH), which posits that people will generally respond with increasing affinity towards increasingly humanlike agents until a certain point. At that point, the degree of human likeness becomes "too much", evoking aversive responses instead of affinity [29]. This change in emotional responding is referred to as the *uncanny valley*. Though its nature and governing mechanisms continue to be of debate (e.g., [3, 8, 10, 17, 31]), there is substantial evidence confirming its existence (e.g., [14, 22, 24, 28, 33]). Moreover, the evidence extends beyond human adults to infants/children [21, 26] and even macaque monkeys [35], suggesting the general phenomenon is relatively pervasive, if not robust.

However, many question whether the uncanny valley presents a serious consideration for human-agent interactions. In particular, as the evidence to-date is primarily based on subjective evaluations, the bearing an agent's appearance may or may not have on a person's *behavior* remains unknown. Specifically, consistent with the UVH, highly humanlike

[1] `http://goo.gl/BxZBe6`
[2] `http://goo.gl/2ttL7m`

Figure 1. Artist Tony Matelli's *Sleepwalker* installation on the Wellesley College campus in Wellesley, MA. Photo credit: Megan Strait.

agents are often described as "eerie", "macabre", "unnerving" (e.g., [18]) – generally less likable than their less humanlike counterparts. But, could an agent's unnerving appearance be so emotionally motivating that it causes a person to act aversively (i.e., avoid interacting)?

In response to this questioning, reactions to Matelli's Sleepwalker would suggest yes. Though unintentional, its appearance nevertheless had significant consequences on how, and moreover, *whether* people would interact with it. For instance, proponents of the statue's removal report using strategies such as modifying its appearance (e.g., dressing it up) to "make him less intimidating", or even taking a different path to avoid the statue entirely. These are two examples – situation modification (dressing up the statue) and situation selection (taking a different path) of ways people regulate a significant emotional response (e.g., fear, apprehension, stress).

While emotions can be useful in certain situations (e.g., a fight or flight response to a dangerous stimulus), the stress elicited by an unnerving agent such as the Sleepwalker might not be particularly appropriate for students' daily lives on the campus surrounding the statue. Hence, regulatory mechanisms can be engaged to help modulate the emotions triggered by such a negative stimulus. According to the process model of emotion regulation (ER; [15]), there are five families of processes that help regulate aspects of the emotion generative cycle ([16]). For example, one can select the situations in which one puts themself (*situation selection*) based on the anticipated emotions resulting from the various contexts (e.g., selecting a different route around campus to avoid a stressful encounter with the Sleepwalker). Alternatively, if a person is already in a certain situation, one can change the emotion-provoking aspects (*situation modification*; e.g., dressing up the statue) or attend to different aspects of the environment (*attentional deployment*; e.g., averting one's gaze from the statue), thereby changing the emotional experience.

The implications of these effects and potential responses are particularly important to human-computer and human-robot interaction. If the Sleepwalker is any indication, an agent's appearance can thus pose particular interference with how and, moreover, *whether* a person interacts with it. Thus, while increasing human likeness has shown promise towards improving interactions, it also remains crucial to gain better understanding of the UVH. Hence, the purpose of the present study was to investigate whether the uncanny valley presents a serious consideration for human-agent interactions. Specifically, we wanted to determine whether highly humanlike robots can be so emotionally motivating that they evoke behavioral *aversion* in human observers.

To address this question, we observed subjects' *behavior* by looking at how often people choose to end encounters with highly humanlike agents versus less humanlike or human agents when given the opportunity, as well as their reasons for doing so. In our attempt to evaluate peoples' aversion to humanlike robots, we employed a modification of the button-press paradigm ([39]) for measuring relevant emotion regulatory mechanisms. Here we presented a series of pictures depicting humans and robots of varying human similarity (low, moderate, and high), with the option to press a button if the subject wished to stop looking at a given picture. In addition to the traditional subjective ratings of the agents' appearances, we collected the percentage of button presses and reaction times (RT) to measure the frequency and rapidity of attempts to end encounters with the various agents. We also recorded eye gaze behavior to index overt *attentional deployment*, as well as physiological indices of emotional responding (corrugator activity, heart rate, and skin conductance).

Hypotheses

Based on evidence that people who attribute negative valuations to a robot also report reduced interest in interacting with it (e.g., [34, 36]), we hypothesized that in order to regulate their emotional response, people would engage in *situation selection* or *modification* by ending encounters (pressing the button) more often and faster in response to highly humanlike robots than less humanlike robots or humans (**Hypothesis 1**). Further – while, based on [4, 27], we expected that greater human likeness might moderately increase interest in *not ending* some encounters (**Hypothesis 2a**) – for those that they terminate, we hypothesized they would report doing so due to being unnerved more so in response to the highly humanlike robots versus the other agents (**Hypothesis 2b**).

In addition, we hypothesized that – when the button press (situation selection/modification) is unavailable – people might engage other ER strategies, such as looking away from the unnverving content (**Hypothesis 3**). We specifically focused on *attentional deployment* as reflected in eye tracking, given that looking away from negative content has been observed in prior work examining other forms of ER (e.g., [6, 38]) and moreover, studies of the UVH using gaze behavior (e.g., [26, 35]) show reduced fixation in response to uncanny agents.

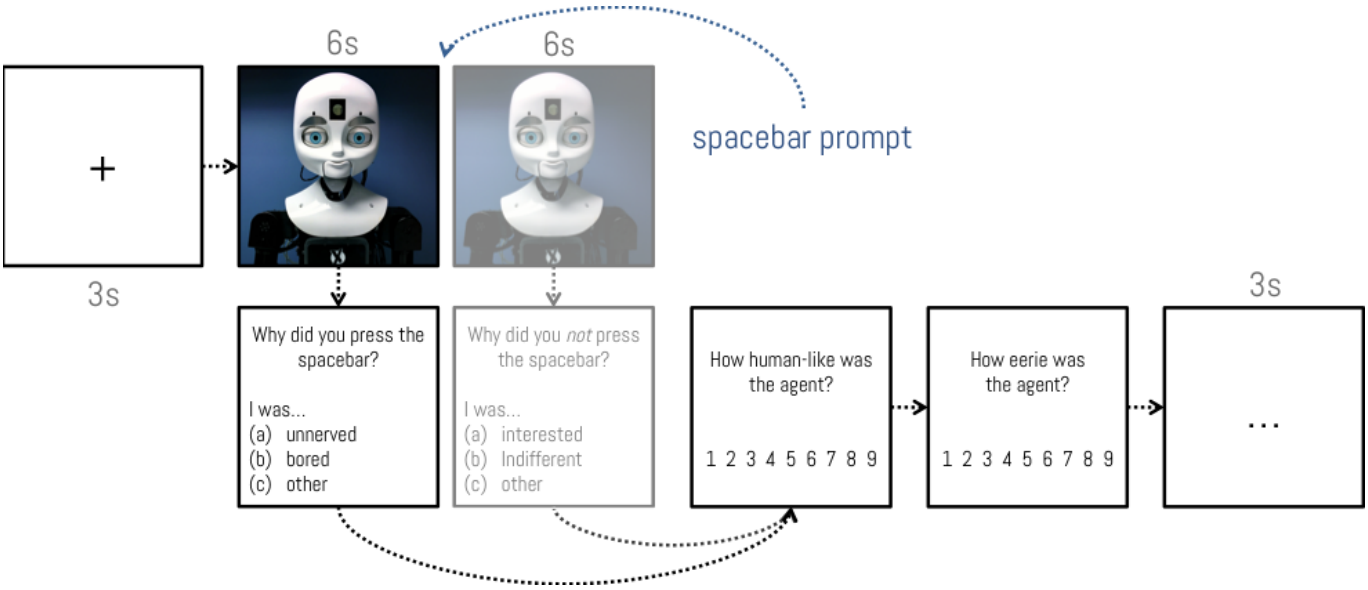

Figure 2. Trial structure: each trial began with a $3s$ fixation point, followed by presentation of an image for at least $6s$. At $6s$, subjects received the prompt: "You may now press the spacebar." If they did not press the spacebar, the image remained for $6s$ more. Otherwise, immediately upon pressing the button, the image was removed and the trial proceeded. After the viewing, subjects completed a series of self-report prompts.

MATERIALS AND METHODS

Based on Mori's UVH, we expected that highly humanlike agents may be so unnerving that people might be motivated to regulate their emotions via situation selection/modification and attentional deployment. To test our predictions, we conducted a fully within-subjects study in which we manipulated **human likeness** of the shown agents (four levels – robots with *low*, *moderate*, and *high* human similarity, and *humans*).

Procedure

Subjects viewed a set of 60 color pictures – each depicting a distinct (robotic or human) agent – selected from a subset of those tested in [37]. The images were obtained from various academic and internet sources and divided into the four *agent categories*. To help mitigate the influence of any one agent, we included 15 instances per category (see Figure 3). The specific agents were selected based on normative ratings showing people perceive these robots as having a low (M=1.80, SD=.13), moderate (M=2.57, SD=.32), or high (M=5.11, SD=.78), with the set of humans as having the highest (M=8.90, SD=.10), degree of human likeness on a nine-point scale (anchored from *machine-like* to *humanlike*).

The above set of 60 images were presented using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA, USA) in an order randomized by subject. Each image was preceeded with a black fixation cross presented in the center of a white screen for $3s$. This was followed by the presentation of the image for between $6s$ to $12s$. Subjects were informed that, after $6s$ of viewing a given image, they could press the spacebar to remove the image from the screen. If the subject did not press the spacebar, the image would remain on the screen for six more seconds (for a maximum viewing duration of $12s$).

The purpose of this $6s$ delay was to impose a minimum encounter duration, during which participants could not modify the situation (via pressing the button). This allowed us to investigate whether people engaged in *attentional deployment* – an emotion regulatory strategy alternative to *situation selection/modification* – when the button pressing was unavailable.

Following the viewing, subjects were cued to select one of three reasons for either pressing (*unnerved*, *bored*, or *other*) or not pressing (*interested*, *indifferent*, or *other*) the spacebar. The choice of these options served to tease apart whether a stimulus creates a negative situation for the subject (being unnerved) or rather, whether a press response indicates boredom. The *other* option was included so that subjects were not presented with a forced choice, but instead could note that they chose to end or not end an encounter with an agent for reasons we did not anticipate. This screen remained present until a response was recorded, at which point, subjects were then prompted to enter the traditional ratings of how humanlike (vs. mechanical) and how eerie (vs. non-arousing) the given agent appeared. The rating screen also remained present until a response was recorded. Lastly, subjects were presented with a white screen showing a black ellipsis for $3s$ to provide a brief break between trials (see Figure 2).

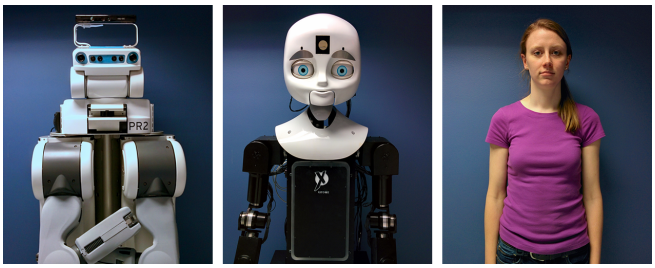

Figure 3. Exemplars of agent categories: robots with *low* and *moderate* human likeness (left and center, respectively), as well as a human (right).

Measures

Subjective ratings of the depicted agents' **humanness** and **eeriness** were used as a manipulation check that our agent categories (*low*, *moderate*, *high*, and *human*) were perceived as having significantly different degrees of human likeness and moreover, that the highly humanlike robots used fall within the proposed uncanny valley. As we used a fully-within subjects design, the ratings were averaged (by subject) across trials within each of the four agent categories.

In addition to the traditional ratings of the agents' appearances, subjects' button press behavior, bilateral eye tracking, and peripheral physiological data were collected using E-Prime 2.0, a Tobii T120 Eye Tracker (Danderyd, Sweden), and MP150 system (Biopac, Goleta, CA, USA) respectively. The details of these recordings and calculation of the dependent variables (DVs) of interest are detailed below.

Behavior

To represent the frequency at which subjects elected to end encounters, the percentage of trials on which subjects pressed the spacebar (**trials terminated**) was calculated within each of the four agent categories. On trials in which the button was pressed, we also calculated the mean **response time** (RT) to represent the speed at which subjects ended encounters.

In addition, we analyzed subjects' reported reasons for their behavior to understand why subjects pressed or did not press the spacebar. On trials in which the button *was pressed*, to determine whether subjects ended encounters due to a negative context associated with the agents' appearance, we calculated the percentage of terminated trials for which subjects reported each of the three responses: **unnerved**, **bored**, or **other**. On trials in which the button was *not* pressed, to determine why subjects elected to view the image for the full duration, we calculated the percentage of these trials for which subjects responded: **interested**, **indifferent**, or **other**.

Lastly, when button pressing was unavailable (during the first $6s$ of viewing), we analyzed subjects' eye gaze behavior to determine whether subjects engaged in *attentional deployment*. Specifically, we computed percentage fixation duration on three areas of interest (AOIs), operationalized as the amount of time subjects spent looking within the pre-defined AOIs for each image relative to total image viewing time. The three areas of interest included: the **agent**, defined as the agent's body and all contiguous, relevant parts (e.g., limbs, hair, clothing); the agent's **head**, defined as the portion of the agent that is clearly separable from a "torso"; and **eyes**, defined as the left/right eye and the area immediately between the two. Fixations within each AOI were identified using the Tobii fixation filter algorithm[3] and were averaged across trials for each AOI within each agent category.

Physiology

In addition to the above, *corrugator electromyography* (EMG), *electrocardiography* (ECG), and *electrodermal activity* (EDA) were sampled as these measures have been found to be sensitive to stimulus valence (e.g., [7, 20]). To index of facial expressive behavior using EMG, two $4mmAg/AgCl$ electrodes were placed in bipolar configuration over the left eye per [12]. Corrugator electromyography was sampled at $1000Hz$ and bandpass-filtered online ($5Hz$ to $3kHz$; $60Hz$ notch filter). Offline, the data were resampled to $400Hz$, rectified and smoothed with a $16Hz$ low pass filter, decimated to $4Hz$, and smoothed with a $1s$ moving average filter.

To measure heart rate, ECG was acquired continuously at $1000Hz$ via two disposable $Ag/AgCl$ electrodes (pregelled with 7% chloride gel and were placed under the left and right collarbones). Offline, the ECG signal was downsampled to $400Hz$ and band pass filtered from 0.5 to $40Hz$. Interbeat interval series were created by identifying R-spikes using automated ANSLAB algorithms. R-spikes that were not detected automatically, thus leading to an erroneously long period between successive R-spikes, were marked for inclusion by hand. Similarly, R-spikes that were identified incorrectly, thus leading to an erroneously short period between successive R-spikes, were removed by hand. Following such artifact correction, the interbeat interval series was converted to HR in beats per minute. HR data were decimated to $10Hz$ and then smoothed with a $1s$ prior moving average filter. Lastly, to index sympathetic activation of the autonomic nervous system, EDA was recorded with DC coupling and constant voltage electrode excitation at $31.25Hz$ via two disposable $1cmAg/AgCl$ electrodes (pregelled with 0.5% chloride isotonic gel and attached to the distal phalanges of the index and middle fingers on the non-dominant hand).

Participants

Sixty-two undergraduates participated in exchange for course credit. All procedures were approved by the Social, Behavioral, and Educational Research Institutional Review Board at Tufts University and subjects provided written, informed consent prior to participating. In anticipation of some loss in data due to issues with the equipment, artifacts, and/or missing values, we chose this sample size in order to achieve at least 50 useable observations in hypothesis testing. Due to equipment failure data were unavailable for two subjects, thus sixty subjects (28 male) with ages ranging from 18 to 28 years (M=19.13, SD=1.48) were included in our final sample.

RESULTS

We hypothesized that, in presenting images depicting agents varying in human likeness, the highly humanlike robots – more so than the less humanlike robots and humans – would be perceived so negatively that subjects would be motivated to regulate their emotions via early termination of their encounters. To test our hypotheses (as well as confirm our underlying assumptions), a repeated-measures ANOVA was conducted on each of the subjective ratings and behavioral DVs with *human likeness* as the IV (see Table 1).[4] The significant results are discussed below, with all post-hoc contrasts reflecting a Bonferroni-Holm correction for multiple comparisons.

[3] `http://www.tobii.com/en/eye-tracking-research/global/library/manuals/`

[4] Due to space constraints, the analyses and results of the peripheral physiological measures are not reported in this paper.

	n	Low	Moderate	High	Human	F	p	η^2
Subjective Ratings								
– *Humanness*	60	.18 (.14)	.33 (.17)	.68 (.17)	.99 (.03)	$F(2.50, 147.94) = 788.85$	$< .01$	.93
– *Eeriness*	60	.29 (.19)	.47 (.17)	.54 (.17)	.06 (.09)	$F(2.47, 145.74) = 219.31$	$< .01$	.82
Press Behavior								
– *Trials Terminated* (%)	60	.52 (.38)	.53 (.37)	.57 (.36)	.57 (.42)	$F(1.81, 106.79) = 1.76$	.17	.02
– *Response Time* (s)	42	1.49 (.80)	1.63 (.85)	1.39 (.51)	1.34 (.50)	$F(2.76, 113.29) = 2.00$	.12	.04
Rationale for...								
– Terminating (%):								
– *unnerved*	42	.26 (.29)	.39 (.32)	.47 (.31)	.04 (.15)	$F(2.50, 102.55) = 36.18$	$< .01$	.46
– *bored*	42	.62 (.32)	.46 (.32)	.39 (.30)	.88 (.21)	$F(3, 123) = 54.85$	$< .01$	.57
– *other*	42	.12 (.19)	.15 (.22)	.13 (.22)	.07 (.15)	$F(2.10, 86.42) = 3.19$	.04	.07
– Viewing (%):								
– *interested*	35	.60 (.34)	.63 (.32)	.70 (.30)	.35 (.36)	$F(2.66, 90.50) = 17.49$	$< .01$	.34
– *indifferent*	35	.39 (.34)	.31 (.34)	.24 (.27)	.60 (.36)	$F(3, 102) = 21.25$	$< .01$	.38
– *other*	35	.01 (.02)	.06 (.12)	.06 (.13)	.05 (.18)	$F(1.77, 60.27) = 1.55$	.22	.04
Gaze Behavior								
– Fixation Duration (%):								
– *agent*	56	.87 (.03)	.84 (.05)	.79 (.05)	.89 (.05)	$F(2.20, 121.09) = 102.91$	$< .01$	.65
– *head*	56	.33 (.05)	.50 (.06)	.57 (.06)	.64 (.07)	$F(2.48, 136.62) = 477.45$	$< .01$	.89
– *eyes*	56	.16 (.05)	.21 (.06)	.21 (.08)	.30 (.11)	$F(1.83, 100.75) = 97.66$	$< .01$	.64

Table 1. Descriptive statistics – number of observations and means (+/-SD) – for the dependent measures as a function of *agent condition* (*low*, *moderate*, *high*, and *human*), as well as inferential statistics (F, p, and *partial* η^2) from testing for main effects of human likeness.

For each ANOVA, the assumption of equal variance was confirmed using Mauchly's test of sphericity or otherwise adjusted. In cases of violation, the degrees of freedom and corresponding p-value reflect either a Greenhouse-Geisser or Huynh-Feldt adjustment as per [13]. We also note that only subjects who provided data in all conditions relevant to each particular test were included (e.g., analyeses of gaze behavior included only those who had non-zero fixations on each of the *agent*, *head*, and *eyes* AOIs). Thus, due to listwise deletion of subjects with missing data, the number of observations (and consequently the degrees of freedom) vary across tests.

Manipulation Check

To confirm the main assumptions of our study design, we first tested subjects' explicit ratings to determine whether our four-level manipulation of *human likeness* – robots with *low*, *moderate*, and *high* human similarity, and *humans* – elicited different attributions of *humanness* and *eeriness*. Specifically, we assumed the four levels would be perceived as having increasing human likeness from *low* (lowest) to *human* (highest). Further, we assumed these categories would elicit differentially negative evaluations, with the greatest eeriness attributed to highly humanlike robots and least to humans.

As expected, a repeated-measures ANOVA across all trials (regardless of whether subjects pressed the button) showed a main effect of *human likeness* on *humanness* ratings (F=788.85, p<.01, η^2=.93). All pairwise comparisons were significantly different (p<.01), confirming that subjects' attributions of human likeness were consistent with those assumed – increasing from agents categorized as *low* (M=.18, SD=.14), *moderate* (M=.33, SD=.17), and *high* (M=.68, SD=.17) in human likeness to *humans* (M=.99, SD=.03).

Similarly, *eeriness* ratings also showed a main effect of *human likeness* (F=219.31, p<.01, η^2=.78). Again, all pairwise contrasts were significant (p<.01) and consistent with assumptions. Specifically, the highly humanlike robots were rated highest (M=.54, SD=.17) and *humans* lowest (M=.06, SD=.09), indicating they were perceived as most and least eerie, respectively. Taken together, these results show that the pictures elicited the expected responding.

Hypothesis Testing

Hypothesis 1

To determine whether people end encounters with highly humanlike robots more frequently and faster than less humanlike robots and humans, we analyzed subjects' press behaviors: % of trials in which the button was pressed (*trials terminated*) and, in those (terminated) trials, the corresponding *response time* (s) to press the button. However, neither measure showed a significant main effect of human likeness (p=.17 and p=.12, respectively). That is, contrary to our expectations, subjects *did not* press the button significantly more frequently or quickly in response to highly humanlike robots than those of low or moderate human likeness and humans.

Hypothesis 2

Despite the non-significant effect of human likeness on termination rates and RTs between agent categories, we expected still that people would show distinct motivations for ending or not ending encounters with highly humanlike agents relative to all others. Specifically, when encounters with highly humanlike robots were *not* terminated, we hypothesized subjects would report doing so out of *interest* (vs. indifference) more so than in response to humans or robots with less human similarity. Conversely, when encounters with highly humanlike robots were indeed *terminated*, we hypothesized subjects would report doing so due to being unnerved (vs. bored).

H2a: Does greater human likeness elicit greater interest? To determine whether increasing human likeness increases

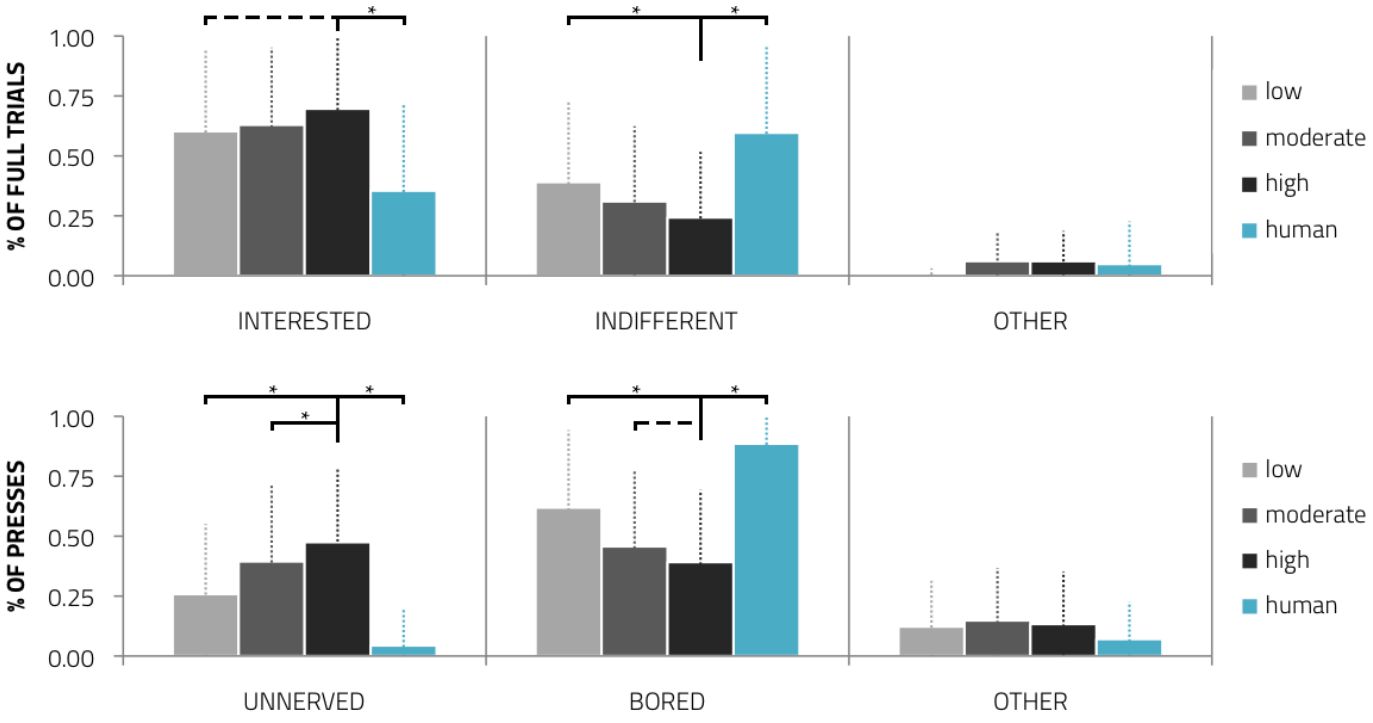

Figure 4. Response frequencies: mean percentage ($+/-SD$) of trials *viewed in full* due to being *interested*, *indifferent*, or *other* (top) and of *terminated trials* due to being *unnerved*, *bored*, or *other* (top) by human likeness category. Asterisks denote significance ($p < .05$) for planned contrasts (between highly humanlike robots and the three other agent conditions), with dashed lines denoting marginal significance ($p < .01$).

subjects' interest in *not* ending encounters with highly humanlike robots, we examined the frequency at which they reported being *interested*, *indifferent*, or *other* on trials that they viewed in full (see Figure 4, top). Repeated-measures ANOVAs showed a main effect of *human likeness* on both *interest* and *indifference* (F=17.49, p<.01, η^2=.34; F=21.25, p<.01, η^2=.38), but no significant effect on the frequency of "other" (p=.22), as reasons for not pressing the button.

As expected, subjects chose to *not* press the button out of indifference less often in response to robots of high (M=.24, SD=.27) versus low (M=.39, SD=.34, p<.01) human likeness. Conversely, they reported marginally greater interest in viewing highly humanlike robots (M=.70, SD=.30) relative to those low in likeness (M=.60, SD=.34, p=.08). However, subjects showed interest/indifference no more or less often in response to the highly humanlike robots than those with moderate human likeness (p=.18; p=.14).

More notably, on these trials where subjects *did not* press the button, they did so because they were more interested in robots – regardless of the degree of human likeness – than humans (M=.35, SD=.36, $p < .01$). Subjects also reported less indifference to robots (similarly, regardless of their human likeness) than humans (M=.60, SD=.36, p<.01).

H2b: Do people end encounters with highly humanlike robots because they are unnerved? To determine whether people ended encounters due to being unnerved more so by the highly humanlike agents relative to less humanlike or human agents, we examined the frequency at which they reported being *unnerved*, *bored*, or *other* on trials that they chose to terminate (see Figure 4, bottom). The analyses showed a main effect of condition on both responses of *unnerved* and *bored* (F=36.18, p<.01, η^2=.46; F=54.85, p<.01, η^2=.57).

As hypothesized, pairwise contrasts revealed subjects chose to press the button because they were unnerved more frequently in response to the highly humanlike robots (M=.47, SD=.31) relative to robots with low (M=.26, SD=.29, p<.01) and moderate human likeness (M=.39, SD=.32, p=.02), as well as humans (M=.04, SD=.15, p<.01).

Conversely, subjects chose to press the button due to boredom less often in response to the highly humanlike robots (M=.39, SD=.30) than to humans (M=.89, SD=.21, p<.01) and robots with low (M=.62, SD=.32, p<.01) and moderate (M=.46, SD=.32, p=.06) human likeness. There was also a main effect of human likeness on presses for reason of "other" (F=3.19, p=.04, η^2=.07); however, no pairwise contrast was significant ($p < .05$), even marginally ($p < .10$).

Although subjects did not terminate encounters with highly humanlike robots any more frequently than those with lesser degrees of human likeness or humans, these results show that their motivations for pressing (or not pressing) the button were nevertheless significantly different with respect to the highly humanlike robots. Importantly, they show that when people do end encounters with highly humanlike robots, they do so because they are unnerved more so than bored.

Hypothesis 3

To determine whether people used attentional deployment to help regulate their emotions *when button pressing was un-*

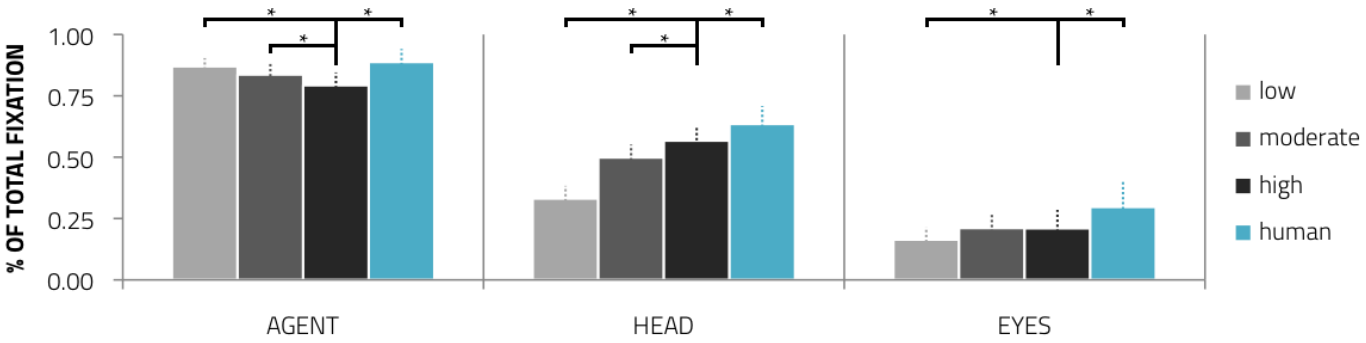

Figure 5. Gaze behavior: mean percent of *total fixation* (+/- one standard deviation) spent fixating on the *agent*, the agent's *head*, and the agent's *eyes* by human likeness category, and their planned contrasts (between highly humanlike robots and the three other agent conditions).

available, we analyzed how participants attended to the images as reflected by the mean percentage fixation duration on three AOIs out of total fixation (see Figure 5). The analyses show a main effect of *human likeness* on all areas of interest: the agent (F=102.91, p<.01, η^2=.65), and the agent's head (F=477.45, p<.01, η^2=.89) and eyes (F=96.67, p<.01, η^2=.64). With respect to the *agent* as a whole, post-hoc contrasts revealed a drastically lower (p<.01) duration spent fixating on the highly humanlike robots (M=.79, SD=.05) relative to the other agents (M_{low}=.87, SD=.03; $M_{moderate}$=.84, SD=.05; and M_{human}=.89, SD=.05).

As to what subjects fixated on within the agent's body, pairwise comparisons of the fixations on the *head* AOI showed significantly different (p<.01) and increasing durations with increasing human likeness: M_{low}=.33, SD=.05; $M_{moderate}$=.50, SD=.06; M_{high}=.57, SD=.06; and M_{human}=.64, SD=.07. Similarly, fixation on the *eyes* increased with increasing human likeness: M_{low}=.16, SD=.05; $M_{moderate}$=.21, SD=.06; M_{high}=.21, SD=.08; and M_{human}=.30, SD=.11. However, the contrast between robots with moderate versus high likeness was not significant.

These results show that increasing the realism of humanlike features (head and eyes) increases attention to such areas. Yet, the difference in overall fixation on highly humanlike robots (relative to the other agents) indicates that subjects use attentional deployment to avert their gaze away from these agents, thus facilitating regulation of the negative context.

DISCUSSION

Contrary to our expectations, subjects did not end encounters with highly humanlike robots any more frequently or faster than encounters with humans or robots with low and moderate human likeness (*H1*). Rather, differences in the use of ER strategies between encounters with highly humanlike versus those with less humanlike or human agents were only evident in their reasons for button pressing and gaze behavior.

Specifically, when subjects were able to *and did* end encounters with highly humanlike robots, they reported doing so more frequently because they were unnerved (*H2b*). For comparison, when subjects ended encounters with humans or robots of low/moderate human likeness, they did so more frequently due to boredom. This indicates that people engage in situation-targeted ER strategies (pressing a button to end an encounter), and with highly humanlike robots, they do so due to the negative context the agents' appearance creates.

In addition, when button-pressing was unavailable (the first 6s of image presentation), we observed evidence of gaze aversion. Specifically, despite subjects' tendency to fixate longer on salient humanlike features (the eyes and face) with increasing human likeness, they showed substantially lower overall fixation on highly humanlike robots relative to the three other agent categories (*H3*). This suggests that people attempt to avoid encounters with such robots, even when they do not have an explicit option to do so. Taken together, these observations show support for our second and third hypotheses, indicating that people are particularly averse to highly humanlike robots (as evidenced by their engagement of both situation-targeted ER and attentional deployment).

On the contrary, when subjects chose *not* to end encounters, human likeness did show a moderately positive effect (*H2a*) as evidenced by a marginal increase in interest and significant reduction in indifference towards robots of high (relative to low) human likeness. However, there was no gain in interest/reduction in indifference relative to the moderate likeness category and in general, subjects showed almost twice as much interest in robots (>60%) – regardless of their appearance – than humans (35%). This suggests the same degree of interest – with less negative consequences – can be achieved using robots of moderate (instead of high) human likeness.

Links to Existing Literature

Mori (1970/2012) hypothesized that people will (1) show increasing affinity towards agents with increasing human likeness up to a certain point – after which, the high degree of human similarity will (2) elicit aversion ([29]).

Prior research has confirmed the first part of Mori's hypothesis, with significant evidence of people responding more positively towards humanlike agents (both virtual and embodied) relative to those with less human likeness (e.g., [9, 30, 40]). Consistent with such findings, subjects here exhibited greater interest and less indifference in their rationale for *not pressing* the button in response to in highly humanlike robots (relative to robots with low and moderate human likeness).

Other work, using overt ratings of agent appearances (e.g., perceived *eeriness*) has shown support of the second part of

Mori's hypothesis with highly humanlike robots rated as more eerie than their less humanlike counterparts (e.g., [22, 24]). Consistent with such work, subjects here rated the highly humanlike agents as being the most eerie and humans the least.

Theoretical Implications

Here we found strong evidence that behavioral aversion was elicited by highly humanlike robots, as reflected by subjects' attempts to regulate their encounters. This included their rationale for ending encounters with highly humanlike agents, the frequency at which subjects reported being unnerved as a function of the human likeness manipulation, and their attentional deployment away from the unnerving agents.

Subjects' engagement of these emotion-regulatory behaviors (the situation-ending button press and gaze aversion) provide additional evidence of the uncanny valley's emotion elicitation. Furthermore, they underscore the negative implications for human-agent interactions which, prior to this investigation, had yet to receive much attention. That is, the highly humanlike robots (more so than the less humanlike and human agents) were so emotionally motivating that subjects averted their gaze from the stimulus and terminated their encounters.

The results also help to shed light on the link between gaze behavior and emotional responding with respect to the uncanny valley. In particular, the correspondance between press behavior, subjective ratings, and gaze lend further support to the conclusions of prior UVH studies using eye tracking data only. Specifically, in cases where overt ratings were unobtainable such as in studying young children [21, 26] and animals [35], this data confirms that lower fixations indicate greater behavioral and subjective aversion. In addition, subjects' increasing fixation on the agent's face and eyes with respect to agent category, suggests that relative human likeness may be reflected by attention to these humanlike characteristics.

Practical Implications

The development of humanlike robots and virtual agents has risen to the forefront of design goals for researchers in both human-robot interaction (HRI) and human-computer interaction, as such agents are increasingly intended for and deployed in social contexts. While there are numerous demonstrations of the benefits of increasing an agent's human likeness in prior work, the current findings suggest that the design of such agents should be done with great care.

Though the stimuli used in the present study were both inoccuous and fleeting, participants nevertheless showed significant aversion to the encounters, however brief. Revisiting the Sleepwalker example: it was a statue, and thereby incapable of actively harming any onlookers. Similarly, the static, image-based stimuli here conferred no possibility of danger to participants and furthermore, were presented for only a matter of seconds. Yet, in both cases, the discomfort caused by the agents' appearances was sufficient to influence peoples' behavior via engagement of emotion-regulatory strategies such as situation selection/modification.

These reactions indicate that there may be a limit to the efficacy of increasing human likeness, and moreover, serious consequences if an agent's appearance goes "too far". Not only do these findings suggest an agent can be "too humanlike", but further, they show that an agent's appearance can significantly affect a person's behavior (such that he/she will avoid the encounter). Considering a more realistic interaction context such as an actual interaction with a physical robot – wherein there may be an actual possibility of danger (real or imagined) – such aversions might only be exacerbated.

Beyond the design of effective social agents, this also holds importance for effective evaluations as well. In the current task, we explicitly instructed people to end the situation if they wanted and by inquiring about the reasons for ending a situation, we further cued them to consider that they might do so if they became upset or bored. But in a actual human-agent interactions (whether in-the-wild or in-the-lab), explicit cues regarding opportunities and reasons for regulating ones emotions are rare. Thus, the situation – wherein a participant is discomforted by an agent but is not given the opportunity to regulate their emotional response – presents a potential confound in investigations that do not consider the adverse effects of an agent's appearance and/or behavior.

Limitations & Future Directions

Our approach to investigating the uncanny valley contributes a novel and simple laboratory task to assess the emotionally-motivated behavioral outcomes that follow from encounters with highly humanlike robots. In particular, the use of behavioral indices to monitor attentional deployment and situation-targeted ER strategies augments traditional methods that typically rely on a small set of self-report measures to study the UVH. This contribution is significant because it uncovers the behavioral impacts associated with the UVH, and furthermore, it sheds light on interpretations of eye tracking data with regard to the uncanny valley. That said, while we are confident that the present study was well-suited to address our primary goals, the approach also has its limitations, which serve to underscore important avenues for future research.

One significant limitation to consider is the use of static, image-based stimuli rather than actual embodied agents. At present, due to the cost and accesibility of physical robots, images are the only practical means of evaluating perceptions across a multitude of real agents. With 15 examples per agent category, broad inferences can be drawn regarding a general theoretical construct (human likeness). Whereas, the alternative (one physical representative agent per category) limits any inferences and their extensibility to those few agents used. However, the image-based stimuli are limiting as well. Specifically, due to their simplicity, the findings hold little bearing on how people might respond to interactive, moving, and/or embodied agents. Based on subjects' aversion despite the harmlessness of the photographed agents (relative the their physically-imposing embodiment), we speculate that the aversive behaviors will only be exacerbated in actual human-robot interactions. However, exactly how the observed effects transfer warrants additional investigation.

In addition, this study's participant sampling was also limited. Specifically, it was fairly homogenous – drawing entirely from undergraduates, most of whom are young women,

at an American university. As such, there may be gender-based and/or cultural differences in perceptions of and emotional responding towards the robots that we did not capture in the present work. For instance, prior research indicates that men and women respond to robots in substantially different manners, with men showing more positive responding to robots than women [19, 32]. Other work has shown evidence that peoples cultural background also has a significant influence on perceptions of robots, with Americans showing the most positive responding [5].

There may also be individual differences amongst participants that further influence how they perceive and respond to humanlike robots. For example, Walters et al. showed that in general, people exhibited greater preference for robots with a more humanlike appearance. However, individual differences showed substantial disagreement with the overall effect: specifically, introverts and participants with lower emotional stability showed greater preference for a more mechanical appearance [40]. More recent findings have shown the similar personality characteristics can predict sensitivity to the uncanny valley [23, 25]. If prevalent within a given sample population, this could potentially bias or skew the results. Thus, subsequent work should draw on broader population to obtain more diverse demographics.

Finally, one question to consider is how to facilate peoples' emotion regulation when the interaction is unavoidable. In the present work we explicitly instructed people to end the situation if they wanted. But in actual interaction contexts, the need to continue interacting with an agent (e.g., to accomplish a collaborative task) might outweigh the motivation to terminate. In this case, it may be of benefit to consider robot-centric strategies to help modulate aversion arising due to its appearance. For instance, the agent's use of gaze aversion (e.g., [1, 2]) or polite speech (e.g., [36]) are two behavioral mechanisms shown to improve perceptions of the agent. Thus, should a person be tasked with interacting with an agent that they find to be "too humanlike", such efforts may have promise towards yielding better emotional outcomes.

CONCLUSIONS

We sought to determine whether the uncanny valley presents a serious consideration for human-agent interactions. Specifically, we investigated whether highly humanlike robots could be so unnerving that they motivate people to avoid them. To do so, we appropriated a simple laboratory task to observe whether subjects engaged in the emotion regulatory strategies, situation selection/modification, by deciding to end encounters with agents of varying human likeness. We also monitored subjects' gaze behavior, as an indication of attentional deployment and further evidence of overt aversion. Our results indicate that people attempt to avoid unnerving encounters with highly humanlike robots (via both situation-targeted strategies and attentional deployment) more so than encounters with less humanlike or human agents. Overall, the current study provides further support of Mori's hypothesis and its relevance to human-robot and human-computer interaction. Should these findings replicate, they provide validation of the valley's emotion elicitation as originally described.

ACKNOWLEDGEMENTS

We are grateful to the research assistants – John Budrow, Alice Chan, Kara Cochran, Melissa Hwang, Chris Martin, Jacob Merrin, Victoria Powell, Rachel Ribakove, and Anoushka Shahane – of the Emotion, Brain, & Behavior Laboratory at Tufts University for helping with data collection and processing; and – Brendan Fleig-Goldstein and Maretta Morovitz – of the Tufts Human-Robot Interaction Laboratory for helping with the literature review. We are also acknowledge Jeffrey L. Birk (now at Columbia University) and Philipp C. Opitz (now at the University of Southern California) for help on the conceptualization and implementation of the study. This research was funded by and conducted at Tufts University.

REFERENCES

1. Andrist, S., Pejsa, T., Mutlu, B., and Gleicher, M. Designing effective gaze mechanisms for virtual agents. In *Proceedings of CHI* (2012), 705–714.
2. Andrist, S., Tan, X. Z., Gleicher, M., and Mutlu, B. Conversational gaze aversion for humanlike robots. In *Proceedings of HRI* (2014), 25–32.
3. Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. Is the uncanny valley an uncanny cliff? In *Proceedings of RO-MAN* (2007), 368–373.
4. Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. My robotic doppelgänger – a critical look at the uncanny valley. In *Proceedings of RO-MAN* (2009), 269–276.
5. Bartneck, C., Suzuki, T., Kanda, T., and Nomura, T. The influence of peoples culture and prior experiences with aibo on their attitude towards robots. *AI & Society 21*, 1-2 (2007), 217–230.
6. Bebko, G., Franconeri, S., Ochsner, K., and Chiao, J. Look before you regulate: differential perceptual strategies underlying expressive suppression and cognitive reappraisal. *Emotion 11* (2011), 732–742.
7. Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. Emotion and motivation i: defensive and appetitive reactions in picture processing. *Emotion 1*, 3 (2001), 276.
8. Brenton, H., Gillies, M., Ballin, D., and Chatting, D. The uncanny valley: does it exist. In *Proceedings of HCII* (2005).
9. Broadbent, E., Kumar, V., Li, X., Sollers, J., Stafford, R. Q., MacDonald, B. A., and Wegner, D. M. Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. *PloS one 8*, 8 (2013), e72589.
10. Burleigh, T. J., Schoenherr, J. R., and Lacroix, G. L. Does the uncanny valley exist? an empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior 29*, 3 (2013), 759–771.
11. Duffy, B. R. Anthropomorphism and the social robot. *Rob. & Aut. Systems 42*, 3 (2003), 177–190.

12. Fridlund, A. J., and Cacioppo, J. T. Guidelines for human electromyographic research. *Psychophysiology 23*, 5 (1986), 567–589.

13. Girden, E. R. *ANOVA: Repeated measures*. No. 84. 1992.

14. Gray, K., and Wegner, D. M. Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition 125*, 1 (2012), 125–130.

15. Gross, J. J. The emerging field of emotion regulation: an integrative review. *Rev. Gen. Psy. 2* (1998), 271–299.

16. Gross, J. J., and Thompson, R. A. Emotion regulation: conceptual foundations. *Handbook of Emotion Regulation 3* (2007), 24.

17. Hanson, D., Olney, A., Prilliman, S., Mathews, E., Zielke, M., Hammons, D., Fernandez, R., and Stephanou, H. Upending the uncanny valley. In *Proceedings of AAAI* (2005), 1728.

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

19. Kuo, I. H., Rabindran, J. M., Broadbent, E., Lee, Y. I., Kerse, N., Stafford, R., and MacDonald, B. A. Age and gender factors in user acceptance of healthcare robots. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, IEEE (2009), 214–219.

20. Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology 30*, 3 (1993), 261–273.

21. Lewkowicz, D. J., and Ghazanfar, A. A. The development of the uncanny valley in infants. *Developmental Psychobiology 54*, 2 (2012), 124–132.

22. MacDorman, K. F. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *Proceedings of CogSci* (2006), 26–29.

23. MacDorman, K. F., and Entezari, S. O. Individual differences predict sensitivity to the uncanny valley. *Interaction Studies* (2015).

24. MacDorman, K. F., Green, R. D., Ho, C.-C., and Koch, C. T. Too real for comfort? uncanny responses to computer generated faces. *Computers in Human Behavior 25*, 3 (2009), 695–710.

25. MacDorman, K. F., Vasudevan, S. K., and Ho, C.-C. Does japan really have robot mania? comparing attitudes by implicit and explicit measures. *AI & society 23*, 4 (2009), 485–510.

26. Matsuda, Y.-T., Okamoto, Y., Ida, M., Okanoya, K., and Myowa-Yamakoshi, M. Infants prefer the faces of strangers or mothers to morphed faces: an uncanny valley between social novelty and familiarity. *Biology Letters 8*, 5 (2012), 725–728.

27. Meltzoff, A. N. 'like me': a foundation for social cognition. *Dev. Sci. 10*, 1 (2007), 126–134.

28. Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception 2*, 1 (2011), 10.

29. Mori, M., MacDorman, K. F., and Kageki, N. The uncanny valley [from the field]. *IEEE Robotics & Automation 19*, 2 (2012), 98–100.

30. Riek, L. D., Rabinowitch, T.-C., Chakrabarti, B., and Robinson, P. How anthropomorphism affects empathy toward robots. In *Proceedings of HRI* (2009), 245–246.

31. Rosenthal-von der Putten, A. M., and Kramer, N. C. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior 36*, 0 (2014), 422–439.

32. Schermerhorn, P., Scheutz, M., and Crowell, C. R. Robot social presence and gender: Do females view robots differently than males? In *Proceedings of HRI*, ACM (2008), 263–270.

33. Seyama, J., and Nagayama, R. S. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence 16*, 4 (2007), 337–351.

34. Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., and Broadbent, E. Mind perception and attitudes towards robots predict use of an eldercare robot. *Int. J. Soc. Rob. 6*, 1 (2014), 17–32.

35. Steckenfinger, S. A., and Ghazanfar, A. A. Monkey visual behavior falls into the uncanny valley. *PNAS 106*, 43 (2009), 18362–18366.

36. Strait, M., Canning, C., and Scheutz, M. Investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. In *Proceedings of HRI* (2014), 479–486.

37. Strait, M., and Scheutz, M. Measuring users' responses to humans, robots, and human-like robots with functional near infrared spectroscopy. In *Proceedings of RO-MAN* (2014), 1128–1133.

38. van Reekum, C., Johnstone, T., Urry, H., Thurow, M., and Schaefer, H. Gaze fixations predict brain activation during the voluntary regulation of picture-induced negative affect. *Neuroimage 36* (2007), 1041–1055.

39. Vujovic, L., Opitz, P. C., Birk, J. L., and Urry, H. L. Cut that's a wrap! Regulating negative emotion by ending emotion-eliciting situations. *Frontiers in Psy. 5* (2014).

40. Walters, M., Syrdal, D., Dautenhahn, K., Te Boekhorst, R., and Koay, K. Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots 24*, 2 (2008), 159–178.