The uncanny advantage of using androids in cognitive and social science research*

Karl F. MacDorman and Hiroshi Ishiguro
School of Informatics, Indiana University / Department of Adaptive Machine Systems, Osaka University

The development of robots that closely resemble human beings can contribute to cognitive research. An android provides an experimental apparatus that has the potential to be controlled more precisely than any human actor. However, preliminary results indicate that only very humanlike devices can elicit the broad range of responses that people typically direct toward each other. Conversely, to build androids capable of emulating human behavior, it is necessary to investigate social activity in detail and to develop models of the cognitive mechanisms that support this activity. Because of the reciprocal relationship between android development and the exploration of social mechanisms, it is necessary to establish the field of android science. Androids could be a key testing ground for social, cognitive, and neuroscientific theories as well as platform for their eventual unification. Nevertheless, subtle flaws in appearance and movement can be more apparent and eerie in very humanlike robots. This uncanny phenomenon may be symptomatic of entities that elicit our model of human other but do not measure up to it. If so, very humanlike robots may provide the best means of pinpointing what kinds of behavior are perceived as human, since deviations from human norms are more obvious in them than in more mechanical-looking robots. In pursuing this line of inquiry, it is essential to identify the mechanisms involved in evaluations of human likeness. One hypothesis is that, by playing on an innate fear of death, an uncanny robot elicits culturally-supported defense responses for coping with death's inevitability. An experiment, which borrows from methods used in terror management research, was performed to test this hypothesis.

Keywords: android science, automatic responses, Masahiro Mori, philosophy of science, uncanny valley

Interaction Studies 7:3 (2006), **297–337**.
ISSN 1572–0373 / E-ISSN 1572–0381

Introduction

An experimental apparatus that is indistinguishable from a human being, at least superficially, has the potential to contribute greatly to an understanding of face-to-face interaction in the social, cognitive, and neurosciences. It would be able to elicit the sorts of responses, including nonverbal and subconscious responses, that people typically direct toward each other. Such a device could be a perfect actor in controlled experiments, permitting scientists to vary precisely the parameters under study. The device would also have the advantage of physical presence, which simulated characters lack. Unfortunately, no such device yet exists, nor will one any time soon; nevertheless, each new generation of robots is coming progressively closer to simulating human beings in appearance, facial expression, and gesture (Minato, Shimada, et al., 2004; MacDorman et al., 2005; Matsui et al., 2005). These robots are also coming down in price and could soon be available for widespread laboratory use.[1]

Very humanlike robots (Figure 1, left) are often referred to as *androids* in the robotics literature to distinguish them from mechanical-looking humanoid robots (Figure 1, right),[2] although robots like Albert Hubo blur the distinction (Figure 3). An android is defined to be "an artificial system designed with the ultimate goal of being indistinguishable from humans in its external ap-

Figure 1. Hiroshi Ishiguro's Intelligent Robotics Laboratory at Osaka University developed the two robots shown above, the android on the left named Repliee R1, a joint effort with Kokoro Co., Ltd., and the humanoid on the right named Eveliee P1, a joint effort with Mitsubishi Heavy Industries.

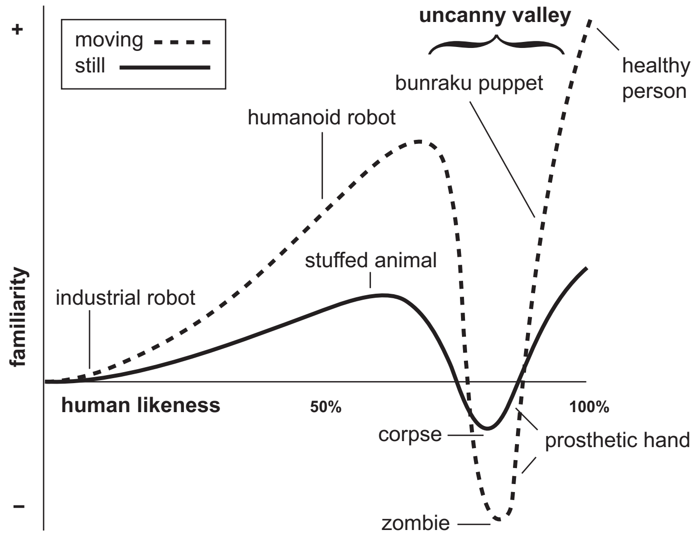

Figure 2. As a robot designer, Mori (1970) graphed what he saw as the relation between human likeness and perceived familiarity: familiarity increases with human likeness until a point is reached at which subtle deviations from human appearance and behavior create an unnerving effect. This he called the *uncanny valley*. According to Mori (1970), movement amplifies the effect.

pearance and behavior" (MacDorman & Ishiguro, 2005). The ability to sustain humanlike relationships with people would be a milestone in the development of androids. Defining androids in terms of a goal and not the current state of the art allows us to distinguish androids from humanoids among present day robots while recognizing that what it means to be an android is very much a moving target for researchers.

In the past, the value of building androids, as opposed to humanoids, was not widely appreciated in the robotics community. In 1970, Masahiro Mori, an influential roboticist, cautioned against building robots that appear too humanlike because they could be eerie or unsettling (Mori, 1970). Some researchers are concerned about the general public's acceptance of androids, which have often been presented ominously in science fiction as human replacements. The media's tendency to sensationalize perceived dangers has the potential not only to undermine funding for android research but for other areas of robotics. Therefore, it is not surprising that in Japan, for example, considerable resources have gone into developing mechanical-looking humanoid robots, such as Honda's Asimo, Sony's Qrio, and Mitsubishi's Wakamaru, while far fewer have been invested in developing androids.

Figure 3. The distinction between android and humanoid is blurred by the robot Albert HUBO, which places the head of Albert Einstein, developed by David Hanson and Hanson Robotics, on the body of the robot HUBO, a robot developed by Jun-Ho Oh and his laboratory at KAIST.

However, an advantage androids do have over other robots is their ability to elicit a person's model of a human other. Indeed, at three meters 70% of study participants mistook the android shown in Figure 5 for a human being during a two-second exposure (Ishiguro, 2005b). One apparent symptom of an android's ability to elicit human-directed responses is a phenomenon Masahiro Mori identified as the *uncanny valley* (Mori, 1970).

Mori predicted that, as robots appear more human, they seem more familiar, until a point is reached at which subtle imperfections give a sensation of strangeness (see Figure 2). He noted that some prosthetic hands are, at first glance, indistinguishable from human hands. However, if you shook one, the lack of soft tissue and cold temperature would give you a shock. The fact that some of these hands can move automatically only increases the sensation of strangeness (as shown by the dashed line in Figure 2). To build a complete android, Mori believed, would multiply this eerie feeling many times over: Machines that appear too lifelike would be unsettling or even frightening

Table 1. Androids provide an experimental apparatus to explore the question of what it means to be human in relation to a number of contemporary debates in cognitive science.

Contemporary debate	Androids are a platform to explore
The relationship between social interaction and internal mechanism	the unification of the social sciences and cognitive psychology
Mind-body problem in cognitive neuroscience	the unification of the psychological and biological
The problem of reductionism in the neurosciences	theories on the creation of artificial consciousness
Connectionism versus modularity in cognitive science	which architectures can produce humanlike response-contingencies
Nature versus nurture	the relative importance of innateness and learning in social interaction
Whether human beings are rational agents	the unification of emotion and reasoning

inasmuch as they resemble figures from nightmares or films about the living dead. Therefore, Mori cautioned robot designers that they should not make the second peak their goal — that is, total human likeness — but rather that they should aspire only to the first peak of humanoid appearance to avoid the risk of their robots falling into the uncanny valley.

The uncanny valley can, however, be seen in a positive light. While many nonbiological phenomena can violate our expectations, the eerie sensation associated with the uncanny valley may be peculiar to the violation of human-directed expectations, which are largely subconscious. If androids are more likely to fall into the uncanny valley than mechanical-looking robots, the reason may be that our brains are processing androids as human. Hence, if androids are capable of eliciting human-directed expectations, human participants can be used to evaluate the human likeness of their behavior to an extent that would be impossible if mechanical-looking robots were used instead. Methodologies from the social and cognitive sciences and ethology can therefore be used to evaluate android performance that were previously used to evaluate human performance (Ishiguro, 2005b). In comparing human-android versus human-human interaction, topics under study include the effects of thinking (MacDorman et al., 2005), lying (Minato, Shimada, Itakura, Lee, & Ishiguro, 2005), and age (Ishiguro, 2005b) on eye contact and gaze. This means that we can use human participants to obtain a more finely-grained analysis of the behavior of androids than is possible with other kinds of robots. This analysis and its subsequent application in improved android designs not only help engineers to build more humanlike androids but also offer insights into human behavior that may be unobtainable by other methods. Androids can act as a testing ground for theories about human interaction and for theories about the role of the brain as a control system in mediating whole-bodied communication. *Therefore, evaluations of cognitive, neuroscientific, and social theories can achieve new levels of sophistication with the help of androids, as opposed to mechanical-looking robots or computer-simulated characters, because people apply a model of a human other to androids, and when that model breaks down, they give the negative evaluation associated with the uncanny valley.* This form of evaluation offers an ecologically valid method of diagnosing faults in models from the social, cognitive, and neurological sciences.

In elaborating this theme, we provide empirical evidence of the uncanny valley and examine a range of possible explanations for the phenomenon. The paper then argues for the usefulness of employing androids in cognitive research for the following reasons: *(1)* The development of androids and the investigation of mechanisms underlying social activity are interdependent; *(2)* android embodiment sets a higher standard for evaluating theories about human behavior than mere robotic embodiment; and *(3)* a synthetic methodology for positing and testing cognitive, behavioral, or neural mechanisms in androids will advance science and engineering.

Cognitive science currently depends on methodologies that often produce dissonant results, especially with reference to issues at its foundations. The methodology of android science offers a new experimental paradigm with the potential to engender breakthroughs in the old cognitive debates (see Table 1). Therefore, there is a case to be made for founding a new, cross-disciplinary field of android science[3] "that integrates the incremental development of robots with the empirical methodologies of the social sciences… [Android science, as broadly construed,] studies the significance of human likeness in human-machine relationships." (MacDorman & Ishiguro, 2006) As an example of android science research, the paper concludes with an empirical inquiry into one explanation of the uncanny valley derived from terror management theory (Pyszczynski, Greenberg, & Solomon, 1999).

Perspectives on the uncanny valley

It seems natural to assume that the more closely robots come to resemble people, the more likely they are to elicit the kinds of responses people direct toward each other. At the same time, subtle flaws in the appearance and movement of very humanlike robots can seem eerie and frightening. Although the uncanny is found in many other contexts (Freud, 1919/2003), we restrict our exploration of the uncanny to human and humanlike forms. As discussed in the introduction, this uncanny phenomenon may be symptomatic of entities that elicit a model of a human other but do not measure up to it. In other words, the android may be uncanny because it fails to respond as predicted by our model of other people and their behavior. If so, a very humanlike robot may provide the best means of finding out what kinds of behavior are perceived as human, because its deviations from human norms are more evident. By setting robot behavior in the context of typical human interaction, robotic results take on ecological validity for human settings and a higher degree of qualitative analysis becomes possible. In pursuing this line of inquiry, it is essential to identify the mechanisms involved in evaluations of human likeness.

The core of Mori's argument for the existence of an uncanny valley in robotics derives from anecdotes about the reactions that a very humanlike prosthetic hand might elicit:

> [R]ecently prosthetic hands have improved greatly, and we cannot distinguish them from real hands at a glance. Some prosthetic hands attempt to simulate veins, muscles, tendons, finger nails, and finger prints, and their color resembles human pigmentation. So maybe the prosthetic arm has achieved a degree of human verisimilitude on par with false teeth. But this kind of prosthetic hand is too real and when we notice it is prosthetic, we have a sense of strangeness. So if we shake the hand, we are surprised by the lack of soft tissue and cold temperature. In this case, there is no longer a sense of familiarity. It is uncanny. (Mori, 1970)

Mori clearly believes that *it is because the hand attempts but fails to replicate human details that it falls into the valley* and not because it is a prosthesis. Indeed, he later argues that a more abstractly designed hand is not uncanny. His explanation of why the realistic hand is uncanny relates to surprise. In more psychological or neurophysiological terms, one might say the realistic appearance of the hand activates tactile expectations or anticipatory reactions that are then violated. However, in a later passage Mori cites the shock and horror that mannequins might elicit, if they started to move. This example does not rely on a cross-modal mismatch, but the violation of expectations mannequins elicit based on our past experience with them. These examples suggest that the cause of the uncanny valley is related to extreme novelty combined with a very humanlike form.

Mori makes a further claim, which is apparent from the graph in Figure 2:

> In mathematical terms, strangeness can be represented by negative familiarity, so the prosthetic hand is at the bottom of the valley. So in this case, the appearance is quite humanlike, but the familiarity is negative. This is the uncanny valley. (Mori, 1970)

This claim is highly controversial. In cognitive psychology and neurophysiology, familiarity is usually characterized as the absence of novelty. Thus, the dependent axis appears to conflate novelty and valence or hedonic value. Wundt (1874) plotted novelty against hedonic value, arguing that hedonic values increase with novelty up to a point at which it declines and, for extreme novelty, becomes negative (cf. Berlyne, 1971). Freud (1919) believed that the uncanny was not to be associated with the unfamiliar; rather, it is something very familiar, but repressed. If there is no contradiction in something being both strange and familiar, the dependent axis would seem misleading.

Can we plot the uncanny valley from experimental data?

Mori's observations about the uncanny valley are based on extrapolation from anecdotal experiences with robots, mannequins, and prosthetic limbs, and they are supported by more recent experience with very humanlike characters in films and video games.[4] There has not yet been an attempt to reproduce the uncanny valley in empirical investigations into human perception using robots of varying degrees of human likeness. As a first step in that direction, an experiment is reported in which human subjects rated still images that included two sets of 11 morphed images on scales of *mechanical versus humanlike, strange versus familiar,* and *eeriness*. In the first set (Figure 4, top), a photograph of the humanoid robot Qrio (left) was morphed into a photograph of an android developed by Hanson Robotics based on the science fiction author Philip K. Dick (center), which in turn was morphed into a photograph of the author himself (right). In the second set (Figure 4, bottom), a photograph of the humanoid robot Eveliee (left) was morphed into a photograph of the android Repliee Q1Expo (center), which in turn was morphed into Repliee's human model (right). To bring out the contrast in eeriness between a prepared and unprepared android, the Philip K. Dick android was left unclothed with hair, scalp, and portions of the skull removed.

Participants

There were 45 Indonesian participants, 37 male and 8 female, of whom 17 were 17 to 20 years old (17 being the age of majority), 18 were 21 to 25, 8 were 26

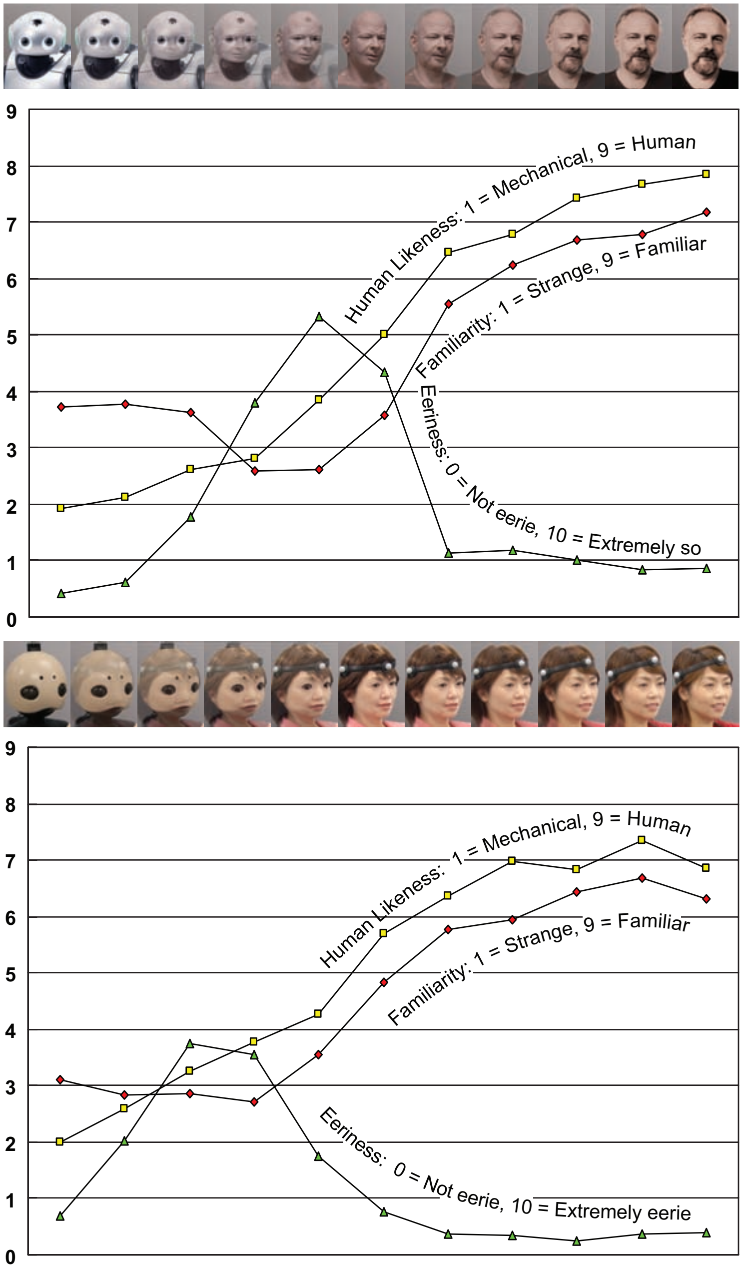

Figure 4. The average ratings of 45 Indonesian participants on scales of human likeness, familiarity, and eeriness are presented for the above figures. The images morph from a mechanical-looking humanoid on the left to an android in the center to a human being on the right (see text).

to 30, and 2 were 51 to 60. The participants were mainly university students, young professionals, and government workers. The prior exposure of the Indonesian participants to humanoid and android robots was minimal, especially relative to people living in Japan where robots receive much greater media coverage. Participants were recruited from an Internet cafe and received two hours of free Internet access.

Procedure

The participants were asked on a computer-based questionnaire, in individual experiments, to rate 31 images presented in random order on a nine-point scale of human likeness and familiarity — that is, a scale ranging from very mechanical to very human and ranging from very strange to very familiar, respectively. The order of presentation of these two ratings questions was also random for each image. Whether the scale went from very mechanical (or very strange) to very humanlike (or very familiar) or the reverse was also set randomly for each question.

The same 31 images were presented again in random order, and the participants were asked, "Which of the following figures gives you an eerie feeling?" The selected images were then presented in random order and the participants were asked to rate how eerie each image was on a ten-point scale, ranging from slightly eerie to extremely eerie. The experiment was conducted solely in *Basa Indonesia*, the official language of Indonesia. An experimenter was on hand to ensure that participants understood what they were meant to do.

Results

The values displayed in Table 2 and Figure 4 are the averages across all participants for human likeness, familiarity, and eeriness. The results show a valley in the *strange versus familiar* scale, where very strange is represented by 1 and very familiar is represented by 9. The trough of the valley occurs in both sets of morphed photographs between the humanoid robot and the android robot. In the region of the valley in both sets, there is a monotonic increase in perceived human likeness moving from the humanoid robot to the android to the human, where very mechanical is represented by 1 and very humanlike is represented by 9. The average standard deviation for the ratings in these two scales is 1.89, the maximum being 2.52 and the minimum 1.04. The peak of the eeriness ratings roughly matches the trough of the familiarity ratings, where not eerie is represented by 0 and maximally eerie is rated by 10. In sum, the experiment reproduced Mori's observations on the uncanny valley with two different sets of morphed photographs.

Table 2. Ratings: Humanlikeness, Familiarity, Eeriness

Image ratio	Humanlike	Familarity	Eeriness
Qrio : Android			
100% : 0%	3.71	1.93	0.42
80% : 20%	3.78	2.11	0.62
60% : 40%	3.62	2.62	1.78
40% : 60%	2.58	2.82	3.80
20% : 80%	2.62	3.84	5.33
0% : 100%	3.58	5.00	4.33
Android : P.K.D.			
80% : 20%	5.56	6.47	1.13
60% : 40%	6.24	6.78	1.18
40% : 60%	6.69	7.42	1.02
20% : 80%	6.78	7.67	0.84
0% : 100%	7.18	7.84	0.87
Eveliee : Repliee			
100% : 0%	1.93	3.71	0.69
80% : 20%	2.11	3.78	2.02
60% : 40%	2.62	3.62	3.76
40% : 60%	2.82	2.58	3.56
20% : 80%	3.84	2.62	1.76
0% : 100%	5.00	3.58	0.76
Repliee : Human			
80% : 20%	6.47	5.56	0.38
60% : 40%	6.78	6.24	0.36
40% : 60%	7.42	6.69	0.24
20% : 80%	7.67	6.78	0.38
0% : 100%	7.84	7.18	0.40

Discussion

Figure 4 constitutes the first attempt to plot the uncanny valley based on data from human participants. The images in the figure morph from a photograph of a mechanical-looking humanoid robot on the left to an android in the center to a human being. For the given images, they reveal an uncanny region, both on the strange-familiar scale and on the eeriness scale. However, David Hanson (2005) argues that a valley of eeriness is not inevitable for a specific range of human likeness. He claims that, across the spectrum of human likeness, it is possible to design androids that are not uncanny. In a follow-up experiment in which intermediate images were designed, adapting more attractive, cartoon-like features, rather than simply morphed, Hanson (2005) eliminates the valley from his results. In addition, Hanson notes that very abstract robots and cosmetically atypical people can be uncanny, although they are far from the posited region of the valley in terms of human likeness.

Do Hanson's results mean that the uncanny valley does not exist? They may suggest that the uncanny valley is not inevitable or that designers with finesse can moderate it in situations that involve still images. Nevertheless, human beings do seem to be highly sensitive to imperfections in near humanlike robots, both in their looks and movements, which is why androids are potentially very useful in studying human perception. While the uncanniness in the morphs may result partially from the fact that the figures are not the product of human design or natural selection, morphing between humanoid robots does not result in uncanniness — or at least not to the same degree. Furthermore, only limited conclusions can be drawn from ratings of still images, which are static, modern inventions appearing after human beings evolved. For instance, cartoon images can be aesthetically pleasing, but if real people could exist with the same proportions, they would be considered freaks. Still images also cannot account for cross-modal mismatches, the effects of movement, and mutual-contingency in interaction. There is no way to evaluate whether a still image is responding as predicted, because they cannot respond at all.

Although Mori asserts that movement increases the effect of the uncanny valley unless it is highly adept, he does not attempt to break down the sense of familiarity into the contributing factors of appearance and behavior (Minato, Shimada, et al., 2004). Goetz et al. (2003) propose to enhance cooperation between a person and a robot by matching the robot's appearance and behavior to the task. The results of an experiment with a "nursebot" named Pearl indicate that the robot's appearance determines what behavior is appropriate. It follows that appearance or behavior that is too humanlike could create unreasonable expectations, producing a negative impression. It is for this reason that the mechanical-looking robot Robovie was intentionally given a mechanical-sounding voice because, ironically, a voice that sounded "too human" would be unnatural (Kanda et al., 2004).

There is also an age-dependent relation (Minato, Shimada, et al., 2004). One-year-old babies were attracted to the Repliee R1 child android; they were unperturbed by even jerky, robotic movements. Children between the ages of three and five, however, were afraid of the android and refused to look at it. Because the baby's model of a human being is underdeveloped, the android was able to pass itself off as human. Adults know that the android is not human, so they do not expect it to fit a human model closely. Young children, however, seem to be in a middle ground of applying a human model to the android but finding it does not match. This is a kind of uncanny valley. These findings indicate the need to build a range of robots, including androids that closely resemble human beings, to examine the complex interplay between appearance, behavior, and the participant's age in giving a comfortable or uncanny impression during interaction.

Explanations of the uncanny valley

There are a number of other possible explanations of the uncanny valley in addition to the idea that androids may be eliciting and violating our expectations about how other people should look and act. These explanations are for the most part untested. In fact, there has been little direct scientific investigation of Mori's uncanny valley hypothesis in the past 35 years. Clearly, there are many qualitatively different ways of deviating from human norms of appearance and movement, some of which are more uncanny than others. In addition, the relation between appearance and behavior in creating a subjective impression of familiarity or human presence has not been well-explored, nor how habituation affects that impression. The uncanny sensation caused by imperfect simulations of human appearance and movement may not be a simple phenomenon, explained by a single mechanism. In part this is because human likeness in both appearance and behavior can vary along many dimensions — each of which can have its own personal, biological, and cultural significance and emotional coloring. With that in mind, we may consider some further explanations:

Expectation violation
As stated above, the more humanlike the robot, the more human-directed (largely subconscious) expectations are elicited. The fact that androids are often incapable of satisfying these expectations may be one reason why we perceive them to be not fully alive. If one person elicits expectations in another, that person elicits contextually-appropriate behavior that, in turn, can produce norm-oriented response (Cowley & MacDorman, in press). Androids violate human expectations about how a person should proceed during interaction. This suggests that some of the peculiarities of interacting with androids may be owing to a failure to model the microbehavior central to the expectational cycle. The elicitation and violation of expectations may occur cross-modally and involve various perceptual and sensorimotor processes. Nevertheless, it seems that not all forms of expectation violation can result in eeriness.

Paradoxes involving personal and human identity

Ramey (2005) proposed a different kind of explanation for the uncanny valley. According to Ramey, an uncanny valley may result from

> any cognitive act that links qualitatively different categories by quantitative metrics that call into question the originally differentiated categories. This effect can be especially pronounced when one of those categories is one's self or one's humanity. From a phenomenological standpoint, humanlike robots may force one to confront one's own being by creating intermediate conceptualizations that are neither human nor robot. (Ramey, 2005)

Unlike Mori, Ramey does not believe the uncanny valley to be a phenomenon that is limited to humanoid robotics.

Evolutionary aesthetics

Another possible explanation for the uncanny valley is that androids are uncanny to the extent that they deviate from norms of physical beauty. A blossoming of research in the past 15 years has shown a biological basis for such norms (Etcoff, 1994, 1999; Johnston & Franklin, 1993; Barber, 1995). Preferences are gender specific and fairly consistent across cultures (Buss, 1989; Symons, 1995). Those deemed attractive by the opposite sex are typically more fertile. Thus, beauty is a potent indicator of potential reproductive success.

The most significant features of beauty are youth, vitality, bilateral symmetry, skin quality, and the proportions of the face and body. Youth is correlated with fertility, especially in women, which peaks around age 20 (Dunson et al., 2002). Youth can be determined from movement alone (Cutting & Kozlowsky, 1997; Barclay et al., 1978) in experiments that display only points of light placed on a person's joints — as can gender, emotion (Dittrich et al., 1996), and identity (Stevenage et al., 1999). Youthful movements tend to be associated with positive emotions (Montepare & Zebrowitz-McArthur, 1988). Such qualities of movement as expressiveness are key indicators of vitality. Not only does expressive behavior enhance attractiveness (Riggio & Friedman, 1986; Sabatelli & Rubin, 1986; Friedman et al., 1988; Depaulo, 1992), but it can reflect fertility. Insofar the skeleto-muscular and nervous systems influence movement, aspects of movement quality could be heritable (Grammer et al., 2003).

Bilateral symmetry is a sign of developmental health. It is correlated with resistance to disease and parasites (Thornhill & Gangstead, 1993), running speed (Manning & Pickup, 1998), elevated sperm count and higher quality sperm (Manning et al., 1998), healthy hormonal levels, and mental wellbeing (Manning et al., 2002). Men generally consider women with a low waist-to-hip ratio (WHR) (Singh, 1993) and mid-level body mass index (BMI) to be more attractive (Tovée et al., 2002), while women prefer men with a higher WHR (Henss, 1995). In women, a low waist-to-hip ratio influences more than attractiveness: women who have large breasts and narrow waists (low WHR) are also more fertile, with 37% higher mean mid-cycle estradiol (E2) (Jasienska et al., 2004). (A woman with 30% higher estradiol is about three times more likely to conceive.) Women also have preferences in the body proportions of men: In one study waist-to-chest ratio accounted for 56% of the variance in male attractiveness (Maisey et al., 1999).

Facial proportions indicate hormonal health (Farkas & Munro, 1987; Thornhill & Gangestad, 1999). Estrogens produce fuller lips and fat deposits in the breasts and hips, while androgens result in a longer and broader lower jaw, wider nose, and more pronounced brow ridges. Men find women more attractive who have larger eyes, fuller lips, and a shorter and narrower lower jaw (Johnston & Franklin, 1993) across various cultures (Perrett et al., 1994; Cunningham et al., 1995), and there is evidence that these features elicit emotional responses in men only (Johnston & Oliver-Rodriguez, 1997). Women prefer male faces that are more masculine when the risk of conception is at its highest during the menstrual cycle (Johnston et al., 2001), and facial dominance in men can predict future military rank (Mueller & Mazur, 1997). Some studies have found consistencies in the norms of facial attractiveness across age, race, and nationality (Perrett et al., 1994). Skin that is free of blemishes indicates the health of the hormonal and immune systems (Symons, 1995; Barber, 1995). It is also more attractive (Fink et al., 2001; Symons, 1995). Unhealthy skin, by contrast, is repellant (Etcoff, 1994; Barber, 1995).

Thus, there are universal norms of beauty that apply to human movement and physical appearance, which are correlated with many physiological qualities including fitness, fertility, and health. Although deviations from these norms are evident along many axes, the decisions that are based on them are often binary — for example, whether to select or reject a particular mate. These judgments have a strong multivalent emotional component whose major axis is attraction–aversion. According to this hypothesis, a robot is uncanny when subconscious processing produces sensations that motivate the beholder to reject it. The uncanny valley may be a symptom of the need to exaggerate differences along the boundaries of such classes as male versus female, fertile versus infertile, child versus adult, living versus dead, human versus nonhuman, and familiar versus unknown.

Rozin's theory of disgust

The natural defense mechanism of disgust may also be related to the uncanny valley. Rozin (1987) and other modern psychologists consider disgust to be an evolved cognitive mechanism to ensure that human beings avoid infection. The more closely another organism is related genetically, the more probable it will be carrying transmittable bacteria and viruses. The reason we perceive certain individuals as attractive is owing to selective pressures on our ancestors, which favor mixing our genes with those of individuals that could maximize the fitness of our progeny. Hence, while organisms with very different genes will not elicit disgust, nor healthy members of our own species, others we may perceive as eerie, if they are diseased or have bad genes. A person with leprosy would be one example.

It is worth noting, however, that Mori refers to the uncanny valley as *bukimi*, which means weird, ominous, or eerie in Japanese. Synonyms of uncanny in English include unfamiliar, eerie, strange, bizarre, abnormal, alien, creepy, spine tingling, inducing goose bumps, freakish, ghastly, and horrible. Disgust is not listed among them. Furthermore, in interviewing and debriefing participants after the experiments reported in this paper, disgust was not mentioned.

But even if eeriness and disgust are different, an explanation that parallels Rozin's analysis of disgust may apply to the uncanny valley. In other words, eeriness may be performing the same function as disgust in Rozin's theory by protecting us from potential exposure to transmittable diseases that can be carried by other members of our species or related species. Mori implies as much in his discussion on death:

> [A] healthy person is at the top of the second peak. And when we die, we fall into the trough of the uncanny valley. Our body becomes cold, our color changes, and movement ceases. Therefore, our impression of death can be explained by the movement from the second peak to the uncanny valley… We might be happy this line is into the still valley of a corpse and not that of the living dead! I think this explains the mystery of the uncanny valley: Why do we humans have such a feeling of strangeness? Is this necessary? I have not yet considered it deeply, but it may be important to our self-preservation. (Mori, 1970)

This leads into the next explanation.

Terror management
One hypothesis is that an uncanny robot elicits an innate fear of death and culturally-supported defenses for coping with death's inevitability. An experiment discussed later in this paper, which borrows from the methods of terror management research, was performed to test this hypothesis. It is easy to see in the following examples how partially disassembled androids could play on subconscious fears of reduction, replacement, and annihilation: (1) A mechanism with a human facade and a mechanical interior plays on our subconscious fear that we are all just soulless machines. (2) Androids in various states of mutilation, decapitation, or disassembly are reminiscent of a battlefield after a conflict and, as such, serve as a reminder of our mortality. (3) Since most androids are copies of actual people, they are *Doppelgaenger* and may elicit a fear of being replaced, on the job, in a relationship, and so on.[5] (4) The jerkiness of an android's movements could be unsettling because it elicits a fear of losing bodily control.

If the uncanny valley is a hodgepodge of different phenomena, it must have a range of causes. Some of these causes may be related more closely to natural selection and neurophysiology, while others seem to depend on conceptual constructs, intentions, and sensorimotor expectations developed during social interaction. Regardless of its cause, heightened sensitivity to any deviations from human looks or movements in very humanlike forms highlights the advantages of using androids relative to other robotic platforms. By bringing these deviations to the fore, androids put us in a better position to correct them.

The interdependence of social science and android development

Osaka University's Intelligent Robotics Laboratory, directed by Hiroshi Ishiguro, has developed four androids in collaboration with Kokoro Co. Ltd.: Repliee R0; Repliee R1 (Figure 1, left), a replica of a Japanese girl at age five; Repliee Q1 also known as Andosan (not pictured, see Matsui et al., 2005); and Repliee Q2 (Figure 5, also known as Uando, which has 42 degrees of freedom). To enhance her aura of human presence, Uando incorporates autonomic responses, such as shifting posture, blinking, and breathing, and reactive behaviors into her verbal and gestural interactions. Since she has been programmed to respond to touch sensors located on her head, shoulders, and arms, if someone, for example, gently taps her shoulder, she will turn and ask "What is it?" However, if someone should try to strike her, she winces, pulls back, and lifts her forearm to protect herself.

Visitors to Ishiguro's laboratory typically cannot help but feel sympathy for Uando in these moments, nor can they be enticed to treat her roughly. This contrasts with demonstrations of the more mechanical-looking humanoid robots' reactions to rough handling. People are inclined to feel far less compassion for Robovie or for Eveliee (Figure 1), a robot based on Mitsubishi's Wakamaru platform. Their "robotic" appearance and behavior cannot elicit the same conscious and subconscious responses that the android does. And, indeed, children who are inclined to roughhouse with Robovie (Kanda et al., 2004) show deference and respect in Uando's presence.

Eliciting interpersonal responses

Certain questions about human beings *can only be answered* by employing androids experimentally. A prime example is the question of how the effects of appearance and behavior interrelate, especially to create an impression of human presence (Minato, Shimada, et al., 2004; Minato, MacDorman, et al., 2004). By employing humanlike robots with varying response-contingencies and facial and bodily proportions, we can evaluate how looks and behavior influence people's experience and responses — not only by questionnaires and other subjective means of evaluation but also through such objective measures as heart rate, respiration, galvanic skin response, speech including prosody, gestures, and the distribution of eye fixations.

Although the range of activity that a human actor can perform is still much broader than the range attainable by an android, androids have already demonstrated distinct advantages. For example, it is nearly impossible for a human actor to hold a static pose. Our bodies are constantly moving, however slightly. In fact, the normal functioning of our visual system depends on this movement. In a two-second Turing test with an android, the importance of autonomic movements in creating an impression of human presence was demonstrated by comparing results for participants who observed a still or moving android. For the static android, only 23% of participants believed the android to be human as compared to 70% for the moving android (Ishiguro, 2005b). Furthermore, the response-contingencies of androids can be precisely fixed for a given experiment, which cannot be said for a human actor. An android can produce, for example, a "sincere" smile on cue, whereas human actors, except those who are highly trained and gifted, tend to produce courtesy smiles that look insincere (e.g., Figure 5 of Thomaz et al., 2005). From the standpoint of testing cognitive theories, it would not make sense to talk about "implementing" a detailed cognitive model of even moderate complexity in a human actor.

Figure 5. Repliee Q2 ("Uando") is controlled by air actuators that provide 43 degrees of freedom. The android can make facial expressions, eye, head, and body movements, and gestures with the arms and hands. Touch sensors with sensitivity to variable pressures are mounted under her clothing and silicone skin. Repliee Q2 uses floor sensors and omnidirectional vision sensors to recognize where people are in order to make eye contact while addressing them during conversation. Repliee's interactions are largely scripted, but she can respond to the content and prosody of her interaction partner by varying what she says and the pitch of her voice.

Moreover, ratios of an android's face and body have the potential to be changed dynamically. Nancy Etcoff (1999) reports that a one millimeter increase in the distance between the eyes in still photographs can make the difference between judgments of attractiveness and ugliness. These sorts of parameters could be manipulated with an android, but not a human actor, to study the importance of such ratios in embodied human forms.

To cite an example from cognitive neuroscience, Dr. Ayse Pinar Saygin is employing Repliee Q2 (Figure 5) in a functional magnetic resonance imaging (fMRI) experiment at the University of California, San Diego. The purpose of the experiment is to study the influence of appearance on human perception of biological motion. In the experiment, the android reaches for a cup. Then the skin of the android is masked or removed revealing the internal electromechanics and precisely the same action is performed. This level of control in movement reproduction cannot be achieved by a human actor, and yet it may be crucial to these kinds of experiments.

During conversation, people from Japan, Europe, and North America tend to make eye contact by fixating on the right eye of their interlocutor. In experiments with an eye motion tracking system, the Eyemark recorder, Minato et al. (2004) confirmed this fixation pattern in graduate students who conversed with a human child or child android. However, the tendency was far less pronounced when participants conversed with a gesturing but mechanical-looking humanoid robot. In a sense, participants were treating the humanoid robot more like an object than a person, since they spent more time looking at parts of the robot's body other than the eyes.

Japanese participants tended to break eye contact when asked questions that required thinking regardless of whether they were interacting with an android or a person (MacDorman et al., 2005). However, they tended to make more eye contact with the android when asked questions that did not require thinking. In Japanese society, too much eye contact between strangers is impolite; and participants appeared to feel less self-conscious about making eye contact with the android than with a human questioner. The study also revealed a difference between Canadian participants, who tended to look up and to the right while thinking, and Japanese participants, who tended to look down, if the questioner was human or an android they believed to be under human control. The downward pattern disappeared for participants who believed the android was under its own control. The importance of this experiment is that it showed, for the first time, that what Japanese participants believe about the mind of the interlocutor influences their gaze behavior while thinking. Regardless of whether a human or android body is sitting in front of the participants, the participants are inclined to show modesty with a downward gaze only if they believe they are interacting with a human being. Minato et al. (2005) found that Japanese participants looked around less when lying to the android as compared to lying to a person. The participants may have thought the android questioner could not detect their deception and, therefore, felt more at ease. It would be far more difficult, if not impossible, to achieve these kinds of results with human actors.

These examples show that to facilitate the most natural and humanlike interaction, we must build androids. *Humanlike appearance and behavior are required to elicit the sorts of responses that people typically direct toward one another.* Since most of our responses are subconscious and inaccessible to introspection, simply "knowing" that a robot can perform humanlike actions is not enough to pull people into normal interactions. The android form immediately tells us what the robot affords (Gibson, 1979) — or ought to — and in a way that a merely humanoid form cannot. The fact that Uando looks and is beginning to act like a Japanese woman sets off a slew of culturally-dependent expectations and responses. For example, it is normal to ask a robot's caretakers for permission to touch it, but a visitor to the lab instead asked his wife, "May I touch her? " because he was concerned how his wife would feel about it. Thus, owing to the android's unique ability to support natural communication, we believe Uando and androids like her constitute a new — but highly familiar — kind of information medium. They can provide a quality of interaction in our daily lives of which ordinary computers — or even humanoid robots — are incapable.

Why do these results run counter to *The Media Equation* (Reeves & Nass, 1996), the notion that people respond to computers and other media with the same social responses that they use in responding to people? Because the kinds of answers one gets have much to do with the way the question is posed. For example, in their work on politeness Reeves and Nass found that human participants would rate the performance of a computer more favorably on 20 of 22 adjectives if the computer being evaluated was the same as the computer requesting the evaluation. These results match the findings of studies in which a person requests a self-evaluation or an evaluation of someone else. Reeves and Nass found similar parallels in attitudes toward computers and people with respect to flattery, judging others and one's self, and teamwork. But in their research, what they evaluate are a participant's linguistic or numeric responses to (verbal) questions *about* an interaction that is already over. *They do not evaluate the interaction itself* — with all of its complex, multimodal dynamics in varying timescales. They do not consider the verbal content of the interaction, the accompanying nonverbal behavior, and the participants' autonomic responses.

To illustrate the importance of nonverbal cues, Dunbar (1996, 2004), Grahe and Bernieri (1999) found that a visual recording of two people planning a trip enabled third parties to make more accurate judgments about their rapport[6] than an auditory recording. A transcript of the words spoken was the least helpful of the three in judging rapport. Since Reeves and Nass typically relied on the typed or written reports of participants about interactions that had already completed, their results may indicate the sociality of this symbolic domain of inquiry more than that of the interaction partner. For example, we are polite to computers insofar as we interact with them symbolically because symbolic modes of interaction are modes of interaction that *Homo sapiens* evolved to commune *with each other*. For the past 150,000 years of human existence, virtually any being that has communicated with symbols has been human. So perhaps human beings are conditioned to see the symbolic as human and to identify the human with the symbolic. In this case it is the symbolic and the human that need to be equated, not computers and people. Even the symbolic is grounded in the dynamics of interpersonal activity. While focusing on symbols as abstracted from those dynamics hides the importance of relationships, individual differences, and embodiment (Cowley & MacDorman, 1995), an examination of whole-bodied communication shows that different media cannot be equated.

Interpreting humans and faces: Specialization and expertise

Human beings have many biomechanical structures that support interaction, including scores of muscles for controlling facial expression, the vocal tract, and hand gestures. They also have neural centers for detecting and interpreting hands and faces. The recognition of faces, facial expressions, and gestures plays a key role in primate communication. Studies based on brain imaging, single cell recordings, face inversion and configurational effects, and recognition deficits (e.g., prosopagnosia) caused by brain injury (Farah et al., 2000) suggest that face recognition is anatomically and functionally specialized (Carmel & Bentin, 2002). Inverted presentation degrades a person's recall of faces much more than for other objects, except in some domains of exceptional expertise (Diamond & Carey, 1986).

Functional magnetic resonance imaging (fMRI) of the brains of human participants has revealed distinct regions within the ventral occipito-temporal (VOT) cortex that respond with high selectivity to faces, objects, or scenes. The face selective region is referred to as the fusiform face area (FFA) (Kanwisher et al., 1997). The FFA has been shown to be more active when the participant perceives the stimulus as a face in near-threshold images (Grill-Spector et al., 2004) or during the bistable oscillation of the Rubin face-vase illusion (Hasson et al., 2001; Andrews et al., 2002). The organization of face-selective regions in the VOT could reflect the kind of detailed discriminations that are required to recognize both faces and their visual attributes (Tarr & Gauthier, 2000). For example, our ability to discriminate between strangers — even when their sensory projections differ only subtly — is well documented, as is our ability to recognize faces from widely differing illuminations and viewing angles.

Homo sapiens may have a genetic predisposition for recognizing faces (Farah et al., 2000), honed by expertise developed over a lifetime. The brain's degree of domain specificity (Spiridon & Kanwisher, 2002) or generality (Gauthier & Logothetis, 2000; Haxby et al., 2001) for recognizing faces and other body parts (Downing et al., 2001) is still contested, as is the relative importance of experience and evolutionary adaptation. Regardless of its origin, however, *human expertise with hands, faces, and facial expressions is automatically applied to expressive machines that closely resemble people,* making androids the most suitable platform for human-robot communication and research.

A synthetic approach to social mechanisms

The use of androids to elicit and mimic human interaction provides a top-down synthetic methodology that could revolutionize the practice of cognitive science. Androids can help us do much more than just discover how people relate to different kinds of robots. Because of their resemblance to people, androids have the potential to contribute to an understanding of human behavior and the roles our brains and bodies play in it. While neuroscience typically takes a bottom-up approach to studying brains by accumulating knowledge about each part and its interrelations (e.g., the hippocampus and its role in spatial memory), simulating human behavior in androids provides a top-down, synthetic methodology for positing and testing response mechanisms at a functionally more abstract level. A top-down approach is called for because our present knowledge of the brain's role in interaction is too limited to assemble the partial results of the analytical approach. By implementing mechanisms to support social interaction in androids, we can elicit interpersonal responses more effectively than we can with mechanical-looking robots. Thus, very humanlike androids can nullify the disruptive effects of appearance, allowing us to focus on human interaction. This allows us to make comparisons with studies that had been performed only on people and to apply methods that have been developed to evaluate human interaction in the social, cognitive, and neurosciences.

Asada et al. (2001) have proposed a synthetic approach to developing and analyzing cognitive models whereby models are implemented in humanoids, their faults are diagnosed, and then the models are improved and reimplemented. Models will need to be revised many times, because there is currently a poor understanding of how the multitude of brain processes integrate and influence behavior.

The psychological literature is replete with methods of evaluating human behavior at various levels, and these methods can be applied both to people and to androids during an interaction. The *Total Turing Test* provides one of the most rigorous methods of evaluating models because a judge would have to find an android's appearance, behavior and, in some forms of the test, even the android's internal workings to be indistinguishable from those of a person (Harnad, 1989, 1991, 1992, 1994, 2000). Turing's original imitation game was devised to evaluate the intelligence of computers under the assumption that mental capacities could be separated from embodiment (Turing, 1950). This begs many questions about the nature of mental capacities and how internal representations are to be grounded in external states of affairs (Harnad, 1990; MacDorman, 1997, 1999). In contrast, the Total Turing Test acknowledges that we have good cause to build androids, because embodiment has proven essential to being human.

As a "pure" test of intelligence, all versions of the imitation game may be flawed, because, for example, a highly intelligent alien species would fail them. French notes that, even in the teletype version of the test, it is possible to pose questions that no computer could answer correctly, if the answer depended on *subcognitive* processes that differ between human beings and machines (French, 1990, 2000). However, a modified Total Turing Test could be used to compare different cognitive models by comparing how true-to-life their responses are, as judged by an independent panel. We have previously referred to this as a *communion game* (Cowley & MacDorman, 1995).

The Total Turing Test is but one method of theory evaluation, but as a general principle, social, cognitive, and neuroscientific models can be embedded in androids and tested in behavioral studies that place androids in the roles of human actors or participants. Once we have controlled for the effects of appearance, this type of study could open the door to a range of new approaches for comparing different theories. In contrast to research in humanoid robotics, the goal of human verisimilitude provides a clear benchmark for scrutinizing models within a research program or for comparing the results of different research teams. The need for a thorough evaluation of both human and android behavior illustrates why the development of androids is beyond the scope of mere engineering: *To make androids humanlike, we must investigate human activity, and to evaluate theories of human activity accurately, we need to implement them in androids, since mechanical-looking humanoid robots do not elicit typical human-directed responses.* This calls for a new field of inquiry that integrates the synthetic approach of robotics with the empirical methodologies of the social and cognitive sciences.

Androids provide not only a testing ground for evaluating cognitive and behavioral theories but also a platform for their eventual unification. Since androids require us to confront issues surrounding both mechanism and behavior, we can no longer view cognition as solely a property of the brain, to be understood at a micro-structural level, nor as socially-definable and separable from biomechanical or sensorimotor constraints. In other words, by implementing neurocognitive mechanisms in androids and then evaluating their interactions with people, androids have the potential to help researchers bridge the gap between cognitive neuroscience and the behavioral sciences, leading to a new way of understanding individuals (Table 1). By taking advantage of this partnership, we hope to find principles underlying the relationship between the human brain and social activity that will apply both to androids and *Homo sapiens*.

Conclusion

Human beings elicit an incredible range of interpersonal responses from each other. Through the processes of natural selection and learning, people are highly sensitized to these responses and to humanlike appearance. Thus, our perceptual systems and our bodies fit together like a hand and glove. Only humanlike appearance and behavior can elicit fully humanlike communication, which is why androids will be one of the most useful platforms for investigating human behavior. Social, cognitive, and neuroscientific models can be implemented in androids that substitute for people in human experiments to compare the effects of their behavior with those of people.

Android implementation can give a more accurate assessment of cognitive models that goes beyond the general benefits of embodiment. At the same time, this assessment gives researchers insight into human behavior, which is essential to making more humanlike androids. Because the development of androids and the investigation of the mechanisms underlying social activity are mutually interdependent, it is imperative to found a new, cross-disciplinary field of *android science* (Ishiguro, 2005b; MacDorman & Ishiguro, 2006). As an avenue for the simultaneous study of mechanism and behavior, android science has the potential to integrate the social, cognitive, and neurosciences.

To these ends, androids will be confronted with circumstances that involve complex, closely coordinated social dynamics, where stable patterns emerge at various spatial and temporal scales, and expectations depend in part on unique histories of interaction. It is out of these social circumstances that androids must construct themselves as social beings, just as human beings have constructed themselves into people.

Notes

* Much appreciation goes to Stephen Cowley, David Hanson, Shoji Itakura, and Takashi Minato for their collaboration; to Heryati Madiapuri for performing experiments; to Christopher Ramey, Dwi Pramono, and Yuli Suliswidiawati for advice on the ethics panel; and to Wayne Christensen, Charles Elkan, Michelle Cowley, Stephen Cowley and the anonymous reviewers for their comments. Sara Kiesler deserves thanks for her suggestion that conflicting evidence for human likeness from a humanlike entity may elicit the same terror management defenses that reminders of death do. Christian Keysers likewise deserves thanks for his suggestion that the uncanny valley may be related to Rozin's theory of disgust. The two experiments on the uncanny valley presented in this paper were performed in the summer of 2005.

1. David Hanson, for example, estimates that the emotionally expressive robot heads currently sold by Hanson Robotics could be mass produced for under US$1000.00. See www.hansonrobotics.com for videos exhibiting their facial animation.

2. Although from their etymology *android* and *humanoid* are synonymous with "resembling man," in robotics *android* generally refers to robots that aspire to a degree of likeness that transcends gross morphology. It is not enough for an android to have a head, two arms, a torso, and perhaps two legs. It must be humanlike down to the look and feel of the skin, teeth, and hair. Its behavior should also approximate human behavior. People should be able to subconsciously respond to the android as human.

3. The term *android science* was proposed by Hiroshi Ishiguro at the IEEE-IFRR Summer School on Human-Robot Interaction, July 23, 2004, Volterra, Italy, then appeared as a paper (Ishiguro, 2005a).

4. The popular media has reported that advances in computer graphics have made animated characters in films and video games realistic to the point of being creepy (Hiltzik, 2001; Seabrook, 2003). Close-ups of Jennifer Garner's face in the video game *Alias* have been described as looking like a "death mask," and close-ups of people in *Resident Evil Outbreak* were "deadeningly weird" (Thompson, 2004). Andy Jones, the animation director of the film *The Final Fantasy,* claimed that figures became eerie and grotesque as they became more realistic: "You start to feel like you're puppeteering a corpse" (Weschler, 2002). Many critics preferred the stylized characters of the film *The Incredibles* to the more humanlike — and more disturbing — characters of *The Polar Express*.

5. After an experiment that included videos of androids, several Indonesian study participants expressed concern about being replaced in the workplace by an android.

6. Rapport denotes a sense of mutual trust, harmony, sympathy, and friendliness, which social robots are often designed to engender (Kanda et al., 2004).

7. For example, in one study "low authoritarian individuals did not derogate attitudinally dissimilar others when mortality was made more salient," while high authoritarian individuals did (Solomon et al., 2000, p. 40), citing an experiment in (Greenberg et al., 1992).

References

Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. *Neuroimage, 17,* 890–901.

Arndt, J., Greenberg, J., Pyszczynski, T., & Solomon, S. (1997). Subliminal presentation of death reminders leads to increased defense of the cultural worldview. *Psychological Science, 8,* 379–385.

Asada, M., MacDorman, K. F., Ishiguro, H., & Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System, 37,* 185–193.

Barber, N. (1995). The evolutionary psychology of physical attractiveness: Sexual selection and human morphology. *Ethology and Sociobiology, 16,* 395–424.

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception Psychophysics, 23, 145–152.*

Becker, E. (1973). *The denial of death.* New York: Free Press.

Berlyne, D. E. (1971). *Aesthetics and psychobiology.* New York: Appleton-Century-Crofts.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12,* 1–49.

Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition, 83,* 1–29.

Cohen, F., Solomon, S., Maxfield, M., Pyszczynski, T., & Greenberg, J. (2004). Fatal attraction: The effects of mortality salience on evaluations of charismatic, task-oriented, and relationship-oriented leaders. *Psychological Science, 15,* 846–851.

Cowley, S., & MacDorman, K. (1995). Simulating conversations: The communion game. *AI & Society, 9,* 116–137.

Cowley, S., & MacDorman, K. (in press). What baboons, babies, and tetris players tell us about interaction: A biosocial view of norm-based social learning. *Connection Science.*

Cunningham, M. R., Roberts, A. R., Wu, C. H., Barbee, A. P., & Druen, P. B. (1995). Their ideas of beauty are, on the whole, the same as ours: Consistency and variability in the cross-cultural perception of female attractiveness. *Journal of Personality and Social Psychology, 68,* 261–279.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9,* 353–356.

DePaulo, B. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111,* 203–243.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115* (2), 107–117.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25,* 727–738.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293,* 2470–2473.

Dunbar, R. (1996). *Grooming, gossip and the evolution of language.* London: Faber & Faber.

Dunbar, R. (2004). *The human story:* A *new history of mankind's evolution.* London: Faber & Faber.

Dunson, D. B., Colombo, B., & Baird, D. D. (2002). Changes with age in the level and duration of fertility in the menstrual cycle. *Human Reproduction, 17,* 1399–1403.

Etcoff, N. L. (1994). Beauty and the beholder. *Nature, 368,* 186–187.

Etcoff, N. L. (1999). *Survival of the prettiest: The science of beauty.* New York: Doubleday.

Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology, 17,* 117–123.

Farkas, L. G., & Munro, I. R. E. (1987). *Anthropometric facial proportions in medicine.* Springfield, Illinois: Charles C. Thomas.

Fink, B., Grammer, K., & Thornhill, R. (2001). Human (Homo sapiens*)* facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology, 115,* 92–99.

French, R. M. (1990). Subcognition and the limits of the Turing Test. *Mind, 99,* 53–65.

French, R. M. (2000). The Turing Test: The first fifty years. *Trends in Cognitive Sciences, 4,* 115–121.

Freud, S. (1919/2003). *The uncanny [das unheimliche]* (D. McLintock, Trans.). New York: Penguin.

Friedman, H. S., Riggio, R. E., & Casella, D. F. (1988). Nonverbal skill, personal charisma, and initial attraction. *Personality and Social Psychology Bulletin, 14* (1), 203–211.

Gauthier, I., & Logothetis, N. K. (2000). Is face recognition not so unique after all? *Cognitive Neuropsychology, 17,* 125–142.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

Goetz, J., Kiesler, S., & Powers, A. (2003, 31 October–2 November). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The twelveth IEEE international workshop on robot and human interactive communication.* Lisbon, Portugal.

Grahe, J. E., & Bernieri, F. J. (1999). The importance of nonverbal cues in judging rapport. *Journal of Nonverbal Behavior, 23,* 253–269.

Grammer, K., Keki, V., Striebel, B., Atzmueller, M., & Fink, B. (2003). Bodies in motion: A window to the soul. In E. Voland & K. Grammer (Eds.), *Evolutionary aesthetics.* Heidelberg: Springer.

Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of the need for self-esteem: A terror management theory. In R. F. Baumeister (Ed.), *Public self and private self* (pp. 189–212). New York: Springer-Verlag.

Greenberg, J., Pyszczynski, T., Solomon, S., Rosenblatt, A., Veeder, M., Kirkland, S., et al. (1990). Evidence for terror management theory ii: The effects of mortality salience reactions to those who threaten or bolster the cultural worldview. *Journal of Personality and Social Psychology, 58,* 308–318.

Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & Breus, M. (1994). Role of consciousness and accessibility of death-related thoughts in mortality salience effects. *Journal of Personality and Social Psychology, 67,* 627–637.

Greenberg, J., Simon, L., Arndt, J., Pyszczynski, T., & Solomon, S. (2000). Proximal and distal defenses in response to reminders of ones mortality: Evidence of a temporal sequence. *Personality and Social Psychology Bulletin, 26,* 91–99.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7* (5), 555–562.

Hanson, D. (2005). Expanding the aesthetic possibilities for humanoid robots. In *IEEE-RAS international conference on humanoid robots.* Tsukuba, Japan.

Harnad, S. (1989). Minds, machines and Searle. *Journal of Experimental and Theoretical Artificial Intelligence, 1,* 5–25.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42,* 335–346.

Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines, 1,* 43–54.

Harnad, S. (1992). The Turing Test is not a trick: Turing indistinguishability is a scientific criterion. *SIGART Bulletin, 3,* 9–10.

Harnad, S. (1994). Levels of functional equivalence in reverse bioengineering: The Darwinian Turing Test for artificial life. *Artificial Life, 1,* 293–301.

Harnad, S. (2000). Minds, machines and Turing: The indistinguishability of indistinguishables. *Journal of Logic, Language, and Information, 9,* 425–445.

Hasson, U., Hendler, T., Ben Bashat, D., & Malach, R. (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain. *Journal of Cognitive Neuroscience, 13,* 744–753.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Henss, R. (1995). Waist-to-hip ratio and attractiveness: Replication and extension. *Personality and Individual Differences, 19,* 479–488.

Hiltzik, M. A. (2001, May 8). Synthetic actors guild. *Los Angeles Times.*

Ishiguro, H. (2005a). Android science (in Japanese). ISCIE Journal, 49, 47–52.

Ishiguro, H. (2005b). Android science: Toward a new cross-disciplinary framework. In *Cogsci-2005 workshop: Toward social mechanisms of android science* (pp. 1–6). Stresa, Italy.

Jasienska, G., Ziomkiewicz, A., Ellison, P., Lipson, S., & Thune, I. (2004). Large breasts and narrow waists indicate high reproductive potential in women. *Proceedings of the Royal Society of London: Biological Sciences, 271* (1545), 1213– 1217.

Johnston, V. S., & Franklin, M. (1993). Is beauty in the eyes of the beholder? *Ethology and Sociobiology, 4,* 183–199.

Johnston, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone mediated adaptive design. *Evolution and Human Behavior, 22,* 251–267.

Johnston, V. S., & Oliver-Rodriguez, J. C. (1997). Facial beauty and the late positive component of event-related potentials. *The Journal of Sex Research, 34,* 188–198.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction, 19* (1–2), 61–84.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

MacDorman, K. (1997). *Symbol grounding: Learning categorical and sensorimotor predictions for coordination in autonomous robots (technical report no. 423).* Cambridge: Computer Laboratory.

MacDorman, K. (1999). Grounding symbols through sensorimotor integration. *Journal of the Robotics Society of Japan, 17,* 20–24.

MacDorman, K. (2005). Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it? In *Cogsci–2005 workshop: Toward social mechanisms of android science* (pp. 106–118). Stresa, Italy.

MacDorman, K., & Ishiguro, H. (2006). Toward social mechanisms of android science: A CogSci 2005 workshop. *Interaction Studies, 7.*

MacDorman, K., Minato, T., Shimada, M., Itakura, S., Cowley, S., & Ishiguro, H. (2005). Assessing human likeness by eye contact in an android testbed. In *Proceedings of the XXVII annual meeting of the cognitive science society.* Stresa, Italy.

Maisey, D. S., Vale, E. L. E., Cornelissen, P. L., & Tovee, M. J. (1999). Characteristics of male attractiveness for women. *The Lancet, 353* (9163), 1500.

Manning, J. T., Gage, A. R., Diver, M. J., Scutt, D., & Fraser, W. D. (2002). Short-term changes in asymmetry and hormones in men. *Evolution and Human Behavior, 23,* 95–102.

Manning, J. T., & Pickup, L. J. (1998). Symmetry and performance in middle distance runners. *International Journal of Sports Medicine,* 205–209.

Manning, J. T., Scutt, D., & Lewis-Jones, D. I. (1998). Developmental stability, ejaculate size and sperm quality in men. *Evolution and Human Behavior,* 273–282.

Matsui, D., Minato, T., MacDorman, K., & Ishiguro, H. (2005). Generating natural motion in an android by mapping human motion. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems.* Edmonton, Canada.

Minato, T., MacDorman, K. F., Shimada, M., Itakura, S., Lee, K., & Ishiguro, H. (2004). Evaluating humanlikeness by comparing responses elicited by an android and a person. In *Proceedings of the second international workshop on man-machine symbiotic systems* (pp. 373–383). Kyoto, Japan.

Minato, T., Shimada, M., Ishiguro, H., & Itakura, S. (2004). Development of an android robot for studying human-robot interaction. In R. Orchard, C. Yang, & M. Ali (Eds.), *Innovations in applied artificial intelligence* (pp. 424–434). Berlin: Springer-Verlag.

Minato, T., Shimada, M., Itakura, S., Lee, K., & Ishiguro, H. (2005). Does gaze reveal the human likeness of an android? In *Proceedings of the fourth IEEE international conference on development and learning* (p. 106–111). Osaka Japan.

Montepare, J., & Zebrowitz-McArthur, L. (1988). Impressions of people created by age-related qualities of their gaits. *Journal of Personality and Social Psychology, 55,* 547–556.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy, 7,* 33–35.

Mueller, U., & Mazur, A. (1997). Facial dominance in Homo sapiens as honest signalling of male quality. *Behavioural Ecology, 8,* 569–579.

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shape and judgement of female attractiveness. *Nature, 386,* 239–242.

Pyszczynski, T., Greenberg, J., & Solomon, S. (1999). A dual-process model of defense against conscious and unconscious death-related thoughts: An extension of terror management theory. *Psychological Review, 106* (4), 835–845.

Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In *IEEE-RAS international conference on humanoid robots.* Tsukuba, Japan.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* New York: Cambridge University Press (CSLI).

Riggio, R. E., & Friedman, H. S. (1986). Impression formation: The role of expressive behavior. *Journal of Personality and Social Psychology, 50* (2), 421–427.

Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of Personality and Social Psychology, 57,* 681–690.

Rozin, P., & Fallon, A. E. (1987). A perspective on disgust. *Psychological Review, 94,* 23–41.

Sabatelli, R., & Rubin, M. (1986). Nonverbal expressiveness and physical attractiveness as mediators of interpersonal perceptions. *Journal of Nonverbal Behavior, 10,* 120–132.

Seabrook, J. (2003, December 1). It came from Hollywood. *New Yorker.*

Simon, L., Greenberg, J., Harmon-Jones, E., Solomon, S., Pyszczynski, T., Arndt, J., et al. (1997). Terror management and cognitive-experiential self-theory: Evidence that terror management occurs in the experiential system. *Journal of Personality and Social Psychology, 72* (5), 1132–1146.

Singh, D. (1993). Body shape and women's attractiveness: The critical role of waist-to-hip ratio. *Human Nature, 4,* 297–321.

Solomon, S., Greenberg, J., & Pyszczynski, T. (1998). Tales from the crypt: The role of death in life. *Zygon: Journal of Religion and Science, 33,* 9–43.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron, 35,* 1157–1165.

Stevenage, S. V., Nixon, M. S., & Vince, K. (1999). Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology, 13,* 513–526.

Symons, D. (1995). Beauty is in the adaptations of the beholder: The evolutionary psychology of female sexual attractiveness. In P. Abramson & S. E. Pinkerton (Eds.), *Sexual nature sexual culture.* Chicago: University of Chicago Press.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3,* 764–769.

Thomaz, A., Berlin, M., & Breazeal, C. (2005). Robot science meets social science: Embodied computational model of social referencing. In *Cogsci–2005 workshop: Toward social mechanisms of android science* (pp. 7–17). Stresa, Italy.

Thompson, C. (2004, June 9). The undead zone: Why realistic graphics make humans look creepy. *Slate.*

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3,* 452–460.

Thornhill, R., & Gangstead, S. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature, 4,* 237–269.

Tov'ee, M. J., Warren, T. T. L., Hancock, P., & Cornelissen, P. L. (2002). Visual cues to female attractiveness: Waveform analysis of body shape. *Proceedings of The Royal Society of London: Biological Sciences, 269* (1506), 2205–2213.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59,* 433–460.

Weschler, L. (2002). Why is this man smiling? Digital animators are closing in on the complex system that makes a face come alive. *Wired Magazine, 10*.
Wundt, W. M. (1874). *Grundzuege der physiologischen psychologie [the main features of physiological psychology]*. Leipzig: Engelmann.

Authors' addresses

Karl F. MacDorman
Indiana University
School of Informatics
IT 487, 535 West Michigan Street
Indianapolis, Indiana 46202 USA
http://www.macdorman.com

Hiroshi Ishiguro
Department of Adaptive Machine Systems
Graduate School of Engineering
Osaka University
Yamada-oka 2–1
Suita, Osaka 565–0871 Japan
http://ed.ams.eng.osaka-u.ac.jp

About the authors

Karl F. MacDorman received a B.A. in computer science from U.C., Berkeley in 1988 and a Ph.D. in machine learning and robotics from Cambridge University in 1996. He is Associate Professor of Informatics at Indiana University (2005–). He was previously Assistant Professor (1997–2000) and Associate Professor (2003–) at Osaka University. He has published more than 40 papers in robotics, machine learning, and cognitive science.

Hiroshi Ishiguro received a B.Eng. and M.Eng. in computer science from Yamanashi University, Japan in 1986 and 1988, respectively. He received a D.Eng. in systems engineering from the Osaka University, Japan in 1991. He is Professor at Osaka University (2002–). He is also Visiting Project Leader at ATR. He was previously Professor at Wakayama University (1999–). His research interests include distributed vision systems, robotics, and android science.

Appendix A: Are humanlike robots uncanny because they remind us of death?

This section attempts to explore one possible explanation of the uncanny valley — that humanlike robots elicit an eerie sensation, because they act as a reminder of our mortality. It attempts to test this hypothesis through the experimental methods used by terror management theory (TMT). TMT studies have correlated subliminal reminders of mortality with a wide range of attitude changes. If an android affects people's attitudes without them knowing it, this raises ethical concerns. If, however, an android is an overt reminder of death, this could impede its future adoption, although people would likely habituate to the effect to some extent. In either case, the looks or movement of the device would need to be enhanced to prevent unwanted effects.

Terror management theory

Like other species *Homo sapiens* are naturally selected for reproductive success, which precludes dying before reproductive maturation. They are therefore motivated to avoid circumstances that lead to death. Yet, unlike other species they are in the potentially terrifying position of knowing that death is inevitable. Inspired by Ernest Becker's *The Denial of Death* (Becker, 1973) and other works, for more than two decades Jeff Greenberg, Tom Pyszczynski, Sheldon Solomon, and their colleagues have been developing a theory concerning how human beings manage their fear of personal extinction (Solomon et al., 1998; Greenberg et al., 1986). The theory has been supported by more than 200 experiments. They posit a dual-process model. Conscious thoughts of death are either suppressed (e.g., by thinking about something else) or their immediate significance is watered down via rationalization (e.g., "My grandmother lived to be 90."; cf. Pyszczynski et al., 1999). Subconscious thoughts of death initiate defense processes that mitigate anxiety concerning the certainty of death by supporting a person's worldview and self-esteem:

> Along with the evolutionary emergence of cognitive abilities that enabled members of our species to comprehend our own mortality, our ancestors developed a solution to the

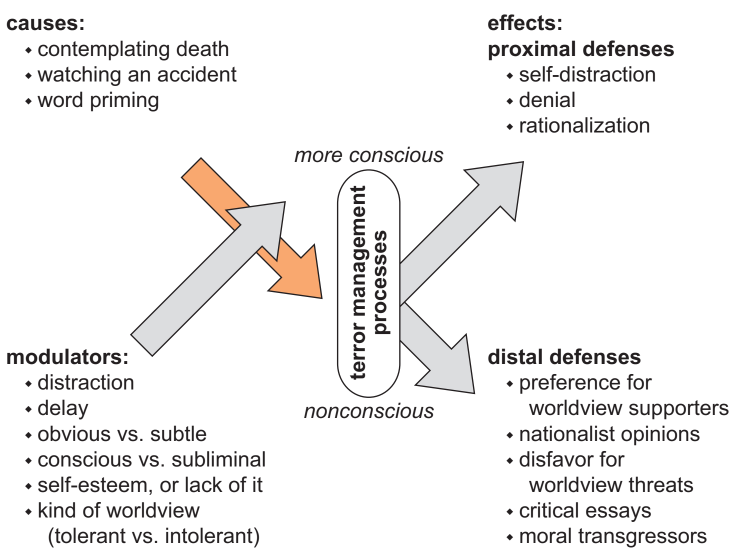

Figure 6. Terror management theory explores the relationship between reminders of death and the defense processes they elicit, including the modulating effects of intervening treatments.

> problem of death in the form of a dual-component cultural anxiety buffer consisting of (a) a cultural worldview — a humanly constructed symbolic conception of reality that imbues life with order, permanence, and stability; a set of standards through which individuals can attain a sense of personal value; and some hope of either literally or symbolically transcending death for those who live up to these standards of value; and (b) self-esteem, which is acquired by believing that one is living up to the standards of value inherent in one's cultural worldview. (Pyszczynsky, 1999, p. 836)

Pyszczynski et al. (1999) contrast the *proximal* terror management defenses elicited by conscious thoughts of death with subliminally-elicited *distal* defenses. Distal defenses may address the threat at a level of abstraction different from that at which it is perceived and understood. Since distal defenses operate outside or at the fringes of consciousness, they need not be rationally connected to the threat and may be best described as *experiential* in nature (Simon et al., 1997).

The *mortality salience hypothesis* predicts that, if having a cultural worldview guards people from anxiety about the inevitability of death (e.g., by giving a literal or symbolic explanation of how death is transcended), those who have been subliminally reminded of death will react more favorably to information that supports their worldview and less favorably to information that undermines it. The hypothesis has been supported by numerous experiments, which have shown, for example, that mortality salience causes people to more strongly prefer essays that praise their country to those that criticize it (Greenberg et al., 1990, 1994, 2000), to prefer charismatic candidates over relationship-oriented candidates (Cohen et al., 2004), and to judge moral transgressors more harshly (Rosenblatt, 1989).

Such distal defenses as worldview protection are activated only after a period of delay, when thoughts of death are conscious (Greenberg, 1994, 2000). However, distal defenses are activated immediately in response to subliminal priming, such as when the word *death* is flashed between the appearance of two other words for an interval too brief to result in one's conscious awareness of it (Arndt et al., 1997). Although a fear of death can produce affective and physiological responses, evidence indicates these responses do not mediate distal defenses; rather distal defenses can occur in their absence (Pyszczynski, Greenberg, & Solomon, 1999).

Appraising human likeness by means of terror management defenses

Do very humanlike stimuli sometimes cause an eerie sensation because they remind us of death and mortality, either consciously or subliminally? For example, an android that is not animated — or not animated like a living person — may look dead. This may remind us, if only subconsciously, of the fact that we too shall die, thus setting in motion defensive mechanisms that influence our attitudes in characteristic ways. If so, we can measure these changes in attitude to explore the terrain of the uncanny valley.

More specifically, the mortality salience hypothesis predicts that subconscious but accessible thoughts of death will provoke distal defenses, resulting in a heightened preference for stimuli that support a person's worldview and an aversion to stimuli that threaten it. If the appearance or behavior of a very humanlike robot, to the extent that it is uncanny, elicits proximal or distal terror management defenses, the effects of these defenses provide a means of quantitatively appraising the human likeness of its appearance and behavior. This then places the focus on the causes of TMT defenses (see Figure 6). So while much research on terror management explores the range of manifestations of terror management defenses (e.g., "Will people who have been reminded of their mortality be more likely to judge moral transgressors harshly? "), the current research assumes that attitude changes that past studies have correlated with mortality salience are valid indicators of worldview defense and begins to consider what stimuli elicit them.

Experiment: Does an uncanny appearance elicit distal defenses?

This experiment was designed to test the hypothesis that an android with uncanny appearance elicits the same distal defenses that reminders of death do. The evaluation criteria are derived from known mortality salience effects in the terror management theory literature: a heightened preference for charismatic politicians relative to relationship-oriented ones (Cohen et al., 2004), and a heightened preference for foreign students who praise a participant's country relative to those who criticize it (Greenberg et al., 1990, 1994, 2000). In addition, mortality salience is gauged by word completion puzzles that are expected to show a participant's preference for death-related word completions, which are indicative of the subconscious activation of death-related associations.

Method

Participants. There were 63 English-speaking participants, 25 male and 38 female, of whom 17 were 16 to 20 years old, 18 were 21 to 25, 9 were 26 to 30, 11 were 31 to 40, and 8 were over 40. Participants were recruited from Zone.com, an online gaming site. The participants were all volunteers and none received remuneration.

Procedure

Instructions. The solicitation for the experiment explained that (1) it involves filling out an online questionnaire; (2) it is for research on a cognitive mechanism that is common to all people; (3) the participant's abilities would not be evaluated; and (4) further details concerning its purpose will be revealed only after the questionnaire has been completed. Potential participants were also told (5) it takes about 10 minutes to complete the questionnaire; (6) it must be completed in order and in one uninterrupted sitting; and (7) they should relax and just give the first answer they think of. Those who agreed to participate were given a link to the questionnaire website.

The website reiterated points 5 to 7 above and summarized the contents of the questionnaire: "You'll be shown some pictures, and you'll be asked some questions to see what you remember about them. Then you will be asked about a couple of excerpts from political speeches and comments made by foreign students. Then you will solve some word puzzles." The wording of the questionnaire was intentionally informal because past studies have found that an informal experimental setting is more conducive to mortality salience

Figure 7. The image on the left is the visual stimulus used for the experimental group. It is the head, neck, and upper torso of an android robot. The eyes are turned up, and there is a gap between the eyes and eye lids, because this part of the android has been powered off and disconnected from the rest of its body. The image on the right is the visual stimulus used for the control group. It depicts an Asian female in her early 20s.

effects (Simon et al., 1997), perhaps because participants tend to base their judgments on gut feelings rather than rational arguments.

Group assignment and stimuli. Participants were randomly assigned with equal probability to either an experimental group or a control group. There were 31 participants in the experimental group and 32 in the control group. Those in the experimental group viewed the uncanny image of an android, while those in the control group viewed the image of a young Asian female (Figure 7). In all other respects, the questionnaire was identical for both the experimental and control group. The participants then viewed in sequence three "filler" images.

Delay. The participants were then asked eight "filler" questions about the images. These questions added a delay before the questions relevant to terror management theory. Past TMT research has found that mortality salience effects appear immediately after subliminal priming on death but only after a delay when death is perceived consciously. Without knowing in advance whether the android would serve as a reminder of death and, if so, whether participants would be conscious of it as such, it was thought prudent to insert a delay.

Worldview-related questions. Participants were next asked to read campaign speeches from two political candidates and to rate on a nine-point scale how well they liked each candidate and how insightful they thought each candidate was. They were then asked which candidate they would vote for. The first speech was *charismatic* and the second was *relationship-oriented* (cf. MacDorman, 2005b for the text). The speeches were loosely paraphrased from a previous study that indicated participants in whom a subconscious fear of death has been elicited are more likely to prefer charismatic leaders (Cohen et al., 2004).

The same five questions were repeated for two foreign students who commented on their experience living in the participant's home country: Participants had to rate on a nine-point scale how well they liked each student, how insightful each student was, and which student they would support if both were running for president of the student government. The first student praised the participants' country, while the second student criticized the participants' country. These questions were inspired by a previous study, which indicated that participants in whom a subconscious fear of death has been elicited are more likely to prefer people who support their worldview (Greenberg et al., 1990, 1994, 2000).

Word completion puzzles. Participants were next given 35 word completion puzzles of the following form:

RELA___G
?

A button under each puzzle read, "Give yourself three seconds to think of the missing letter with a ? under it, and then click here." After clicking the button, the puzzle vanished and several choices appeared, among which the participant could select only one. In the above puzzle, for example, a participant might select *X* to signify *relaxing*:

○ T ⦿ X ○ Y ○ other / don't know

The participants were then taken to the next question.

Following the TMT literature, participants were taken to a set of word puzzles. Dispersed among this set of 35 puzzles were 7 that allowed participants to choose among word completions, one of which was related to death. These puzzles were intended to detect a subconscious activation of death-related concepts. The puzzles in the questionnaire are listed below with italics denoting the death-related option: COFF−−: *coffin,* coffee; SK−−L: skill, *skull*; MUR−−R: murmur, *murder*; GRA−−: grace, grade, grate, *grave*, graze; BUR−E−: burden, burger, *buried*, burned/burner, burped, burred; −EAD: bead, *dead*, head, lead, mead, read; STI−−: stick, *stiff*, still/stile/stilt, stink/stint/sting. A further 7 questions were intended to detect a subconscious activation of concepts that are roughly synonymous with the uncanny.

Suspicion and qualitative remarks. Finally, the participants were asked whether they had any difficulty completing the questionnaire; whether they had any suspicion concerning what the questionnaire was about; and what their impression was of the four images shown at the beginning. Six participants in the experimental group were selected for further questions concerning their impression of the uncanny image of the android. The participants were finally debriefed concerning the purpose of the experiment.

Results

Worldview-related questions. On average, the results show a consistent preference for worldview supporters and against worldview threats in the experimental group (see Table 3 and Figure 8). The experimental group rated the charismatic political candidate nearly a point higher for likeability (+0.93) and insight (+0.80) and rated the relationship-oriented candidate lower on likeability (−0.31) and insight (−0.66). The experimental group rated the foreign student who praised the participants' country higher for likeability (+0.30) and insight (+0.19) and rated the one who criticized it lower on likeability (−0.84) and insight (−0.30). (For preference questions, 1 = strongly negative, 5 = neutral, and 9 = strongly positive.)

The charismatic candidate lost by 7 votes in the experimental group (12 to 19) but by 16 votes in the control group (8 to 24), more than double the margin. The praising and critical foreign students tied in the control group (16 to 16), while the praising foreign student won by a 13 vote margin in the experimental group (22 to 9). (For voting questions, 1 = worldview-supportive candidate and 2 = worldview-critical candidate.) The average standard deviation for questions 1 to 4 and 6 to 9 was 2.00 and for questions 5 and 10 it was 0.48. Student's *t*-test (two tails, heteroscedastic) showed statistical significance overall ($t = 2.17$, degrees of freedom = 61, $p = 0.0348$), as did the Mann-Whitney *U* Test ($n1 = 32$, $n2 = 31$, $U = 658.5$, *P*, two-tailed = 0.024724).

Word completion puzzles. Among the 28 participants in the experimental group, there were 49 death-related word completions as compared to 38 among the 31 participants in the control group, 85 uncanny-related word completions in the experimental group as opposed

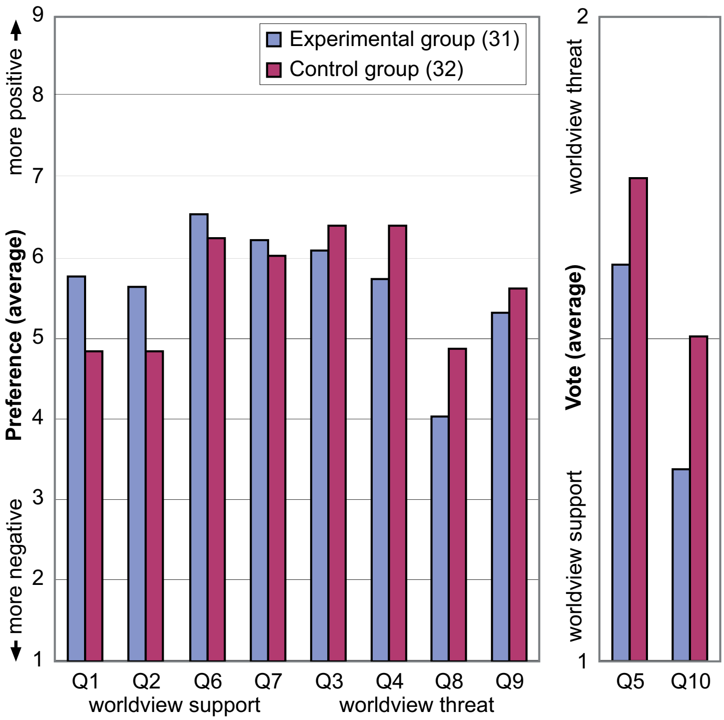

Figure 8. The experimental group showed a stronger preference for the charismatic candidate (questions 1 and 2) and less affinity for the relationship-oriented candidate (question 3 and 4) relative to the control group. They were also less likely to vote for the relationship-oriented candidate (question 5). The experimental group likewise showed more of a preference for the foreign student who praised their home country (questions 6 and 7) than the one who criticized it (questions 8 and 9) relative to the control group, and they voted for the praising student by a wide margin (question 10). Questions 1 through 5 reproduced part of the results of Cohen et al. (2004) and questions 6 to 10 reproduced part of the results of Greenberg et al. (1990, 1994, 2000) but using an android as the experimental stimulus in both cases.

Table 3. Worldview: Average values

Question	Experimental	Control
1	5.77	4.84
2	5.65	4.84
3	6.10	6.41
4	5.74	6.41
5	1.61	1.75
6	6.55	6.25
7	6.23	6.03
8	4.03	4.88
9	5.32	5.63
10	1.29	1.50)

Table 4. Word completion: Totals

Type	Experimental	Control
Participants	28	31
Death	49	38
Uncanny	85	66
Combined	134	104

Table 5. Word completion: Average values

Type	Experimental	Control
Death	1.75	1.23
Uncanny	3.04	2.13
Combined	4.79	3.35

to 66 in the control group, and 134 combined death and uncanny-related word completions as opposed to 104 in the control group (see Table 4). On average the experimental group had more death-related and uncanny-related word completions than the control group (see Table 4).

Student's t-test (two tails, heteroscedastic) and the Mann-Witney U test showed statistical significance for uncanny-related questions ($t=2.59$, degrees of freedom $=57$, $p=0.0132$; $n1=31$, $n2=28$, $U=585.5$, P, two-tailed $=0.020424$), but not for death ($t=1.71$, degrees of freedom $=57$, $p=0.0963$; $n1=31$, $n2=28$, $U=537.0$, P, two-tailed $=0.119108$) related questions. For combined death and uncanny-related questions, the statistical significance was the highest ($t=2.92$, degrees of freedom $=57$, $p=0.00542$; $n1=31$, $n2=28$, $U=629.5$, P, two tailed $=0.002686$).

Discussion

The effects of distal terror management defenses may not be such a reliable indicator of the degree of mortality salience of a given stimulus. The same stimulus will affect the attitudes of individuals differently.[7] In addition, proximal and distal defenses have varying response strengths depending on whether the stimulus is perceived in focal or fringe consciousness or subliminally (Figure 9(a)). The same stimulus can produce varying effects owing to delay, some of which will be too weak to detect owing to the high degree of variance in the data (Figure 9(b)).

A more fundamental concern relates to affect. The eerie sensation identified with the uncanny valley may be characterized as affective, although it seems difficult to identify it with one or more primary emotions like disgust. Terror management studies have indicated that affect does not mediate distal defenses (Pyszczynski, Greenberg, & Solomon, 1999), and it seems that subliminal priming of death-related words does not create an eerie sensation. That leaves the question of what subconscious processes are actually underpinning the

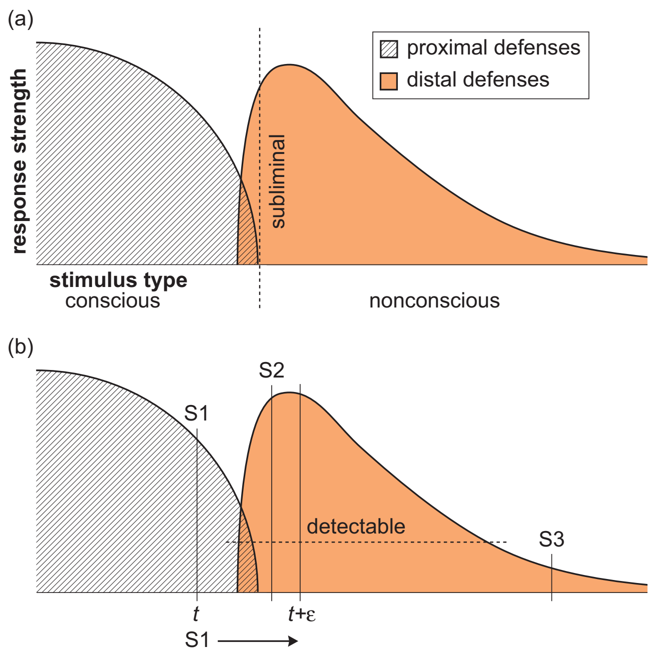

Figure 9. (a) Stimuli may be perceived in focal or fringe consciousness or subliminally. This can produce proximal or distal defenses with varying response strengths. (The curves in the figure are given only for the sake of example; their actual shapes are unknown.) (b) Different stimuli may elicit different kinds of defenses (S1 and S2) as may the same stimulus (S1) at different times (*t*,*t*+[epsilon]) owing to the effects of delay. In addition, for a given sample size, the effects of some stimuli may be impossible to measure owing to variance in the data (S3).

eerie sensation. The uncanny android still seems to be a reminder of death, but for many participants a conscious reminder, in which case the distal (i.e., subconscious) defenses that showed up in the experimental results may have occurred owing to delay.

This experiment investigated the hypothesis that an uncanny-looking android may be uncanny because it elicits a fear of death and attempted to verify this hypothesis with questions designed to measure one distal terror management defense, namely worldview protection. The results are favorable. On average, the group exposed to an image of an uncanny robot consistently preferred information sources that supported their worldview more relative to the control group. The results, however, only apply to one particular stimulus, so it is necessary to ascertain whether they generalize across uncanny stimuli and, in particular, to uncanny movement in a robot that otherwise looks human and natural.