Expanding the Aesthetic Possibilities for Humanoid Robots

David Hanson

Abstract

Although the uncanny exists, the inherent, unavoidable dip (or valley) may be an illusion. Extremely abstract robots can be uncanny if the aesthetic is off, as can cosmetically atypical humans. Thus, the uncanny occupies a continuum ranging from the abstract to the real, although norms of acceptability may narrow as one approaches human likeness. However, if the aesthetic is right, any level of realism or abstraction can be appealing. If so, then avoiding or creating an uncanny effect just depends on the quality of the aesthetic design, regardless of the level of realism.

1. Introduction

In Mary Shelley's excellent novel Frankenstein, the synthetic creature suffers the rejection of humans simply because the creature looks terrible. With skin that is papery and horrid, and eyes that are flaccid and rheumy, he wears many direct emblems of mortality. If only his designer had been more attentive to his appearance.

Imagine if Victor Frankenstein had provided his creature with nice skin, a warmly expressive smile, and big attractive eyes; we may expect that the story would have been much less tragic. Such design modifications would not have made the thing look more human per se, just closer to the aesthetic template for human beauty and good health. Beauty and good health are no more realistic than the morbidity exhibited by the gravely ill or injured. The main factor in the monster's repulsive demeanor, then, is not his level of realism, but bad design.

However, this notion is not consistent with the theory of the uncanny valley—a central guiding principle in robot design for thirty five years.

First posited by roboticist Masahiro Mori [Mori, 1970], the uncanny valley theory holds that the level of realism will directly determine how eerie (uncanny) a humanlike depiction will be. The theory contends that cartoonish depictions of humans are inherently appealing, as are perfectly realistic depictions, but that any depiction in-between will be inherently eerie or uncanny (see figure 1). The "valley" refers to that purportedly unavoidably disturbing region in the middle. Mori contended that the valley existed for static as well as dynamic humanlike figures. He also contended that one should avoid making robots that might land in the valley, out of fear that such robots would be rejected by people.

Although much debated, the theory has not to date been seriously tested or challenged [Hanson et al, 2005]. Out of deference to the theory, most robots are designed to be extremely abstract [Breazeal, 2002; Duffy, 2002; Fong et al, 2003; Caporael, 1990], while a few have been as close to realistic as possible [Hara et al, 1998; Hanson et al 2001, Hanson et al 2005; Menzel and D'Aluisio, 2000]. In between there exists an almost entirely unexplored territory of intermediate designs [Hanson et al 2005].

Many anecdotal examples, however, imply that humans' reactions to an anthropomorphic depiction are more strongly related to good or bad design than to level of realism. Throughout art history, anthropomorphic depictions are considered beautiful independent of their level of realism. For example, some of our most fetishized icons such as the Mona Lisa or Barbie, look almost but not quite real. Likewise, humanlike characters in the paintings of Norman Rockwell, sculptures of Rodin, and the Princess Fionna character of *Shrek*, do not seem eerie in spite of being unmistakably not quite human. And yet, according to the theory, they should be entrenched in the uncanny valley.

Conversely, extremely abstract figures, and perfectly real figures may be found uncanny. For example, real humans may be uncanny, such as people with cosmetically atypical physiologies, psychotic individuals, or gravely ill or injured humans. On the low end of realism, extremely abstract robots may be uncanny, such as MIT's Lazlo (see figure 1) or the University of Reading's Morgui (figure 10).

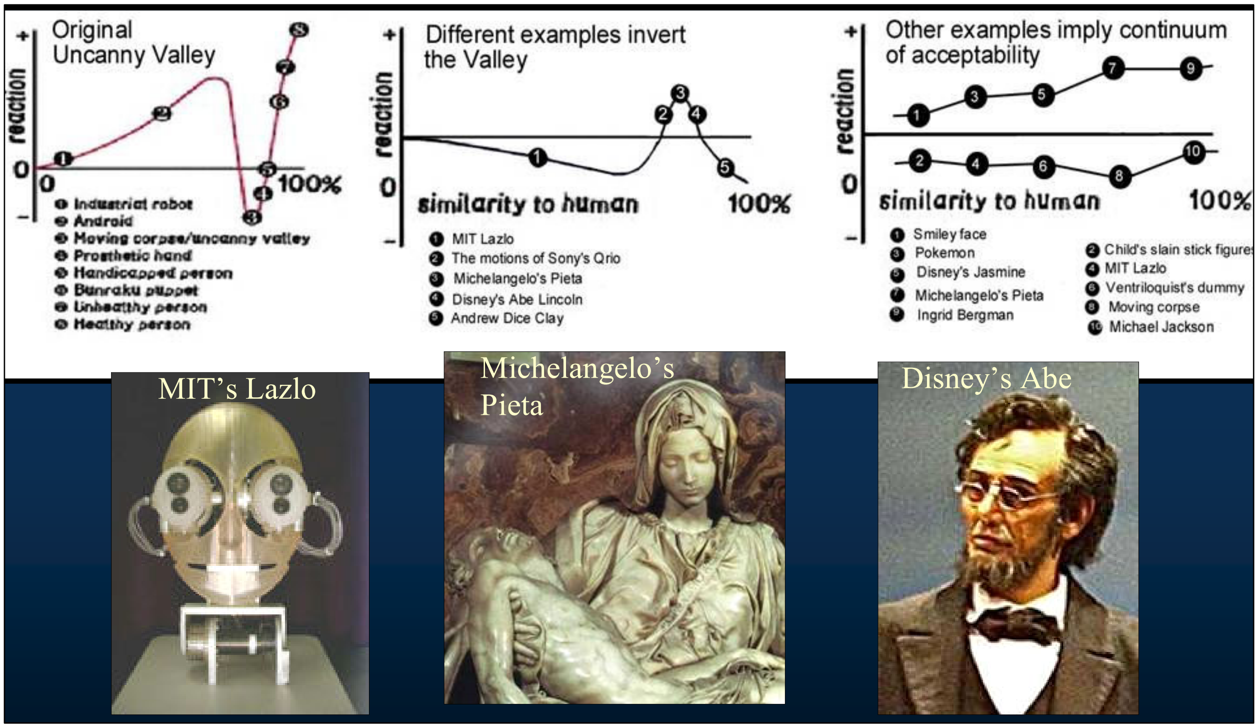

Figure 1. a. Mori's uncanny valley, **b.** inversion of the valley, **c.** dense possibilities with no valley.

These examples seem to turn the uncanny valley graph upside down (figure 1.b). Other well-chosen examples will populate a continuum of positive reaction from abstract to extremely real; while alternate examples demonstrate a continuum of negative reaction across the dimension of realism. These examples strongly imply that realism may not be the determining factor in human repulsion from an anthropomorphic object.

Given these examples, it appears that the uncanny valley theory may not be an accurate characterization of anthropomorphic aesthetic possibility. To really challenge the theory, though, more rigorous testing has been required. This paper describes a series of formal tests that attempt to pose such a challenge to the theory of the uncanny valley. The results of these tests strongly contravene the uncanny valley theory.

The paper then proposes alternative theories.

2. Background

For millennia, artists have stumbled in and out of successful resonance with human perception of beauty. Artistic representations of the human form have proliferated into an incredible diversity, from highly abstract to highly realistic, stunningly beautiful to the outrageously grotesque. Art sometimes makes images that are very eerie. In general, however, artistic representations gravitate (like the human eye [Etcoff, 2000]) towards beauty. The beautiful tend to become the most highly celebrated artworks. In the hands of artists, pigment was transmuted into the Venus de Milo, marble into the David, and animation into Sleeping Beauty.

In recent years, neuroscientists and evolutionary psychologists have found strong evidence that our perception of beauty and ugliness is genetically stamped, deeply into our nervous system [Rhodes and Zebrowitz, 2002]. Abundant evidence supports the contention that people's aesthetic tastes for appearance have been highly refined by evolutionary pressures, into universal, neural-templates that filter distinctly for beauty [Etcoff, 2000; Cunningham et al 2002], for ill health and danger [Darwin and Ekman, 1872/1998; La Bar et al 2003; Etcoff, 2000; Kesler-West et al, 2001], and for "things we are or are not accustomed to" [Dion, 2002].

These neural-templates are probably the attractors that artists work with, and they represent the ultimate obstacle-course for social robot designers as well. Any "uncanny" perceptual phenomenon would, by necessity, be a consequence of these neural systems.

Some of these tastes for humanlike appearance appear mutable, being cultural or otherwise acquired [Dion, 2002; Reiman et al 2000, Cunningham et al, 2002]. Others appear to be *im*mutable, being hardwired into our nervous systems [Etcoff 2001]. If the uncanny valley phenomenon is real, is it immutably so? Or can we grow accustomed to nearly-human looking robots?

The literature on human facial attraction clearly shows that there exist multiple dimensions to humans' reaction to humanlike appearance [Zebrowitz and Rhodes, 2002; Cunningham et al 2002]. Understanding of these dimensions, their dynamics, and the limits of aesthetic possibility is in its infancy. Only by fearlessly exploration can we secure an understanding of the true limits of human-robot design.

The stakes are high. The introduction of new technologies into the public consciousness is a notoriously precarious process, and first impressions can echo for decades or longer. Roboticists have been understandably cautious to make a good first impression [Breazeal, 2002; Duffy, 2002; Fong et al 2003]. On the other hand, for social robots to reach their potential, we must enable creativity over the widest possible range of appearance. Artificial restrictions will prevent social robots from flowering and flourishing fully.

The theory of Bukimi no tani [the uncanny valley] predicts that the figures will be inherently eerie somewhere between cartoons and perfectly real [Mori, 1970]. The original theory further contends that the uncanny valley holds for static figures, but will be stronger for moving figures.

Later speculation considered that is will be important that a figure show the right level of intelligent response to its environment and to social cues to avoid uncanny effects, in effect managing people's expectations as they interact with robots [Breazeal, 2002; Scasselatti, 2005; Fong et al, 2003]. The theory, however, does not effectively define the dimensions of realism, and its examples seem to widely vary in their aesthetics characteristics in ways that are not well defined. For example, a "walking corpse" is ranked as less realistic than a "bunraku puppet" (see figure 1-a).

As a first order of business in our formal challenge of the theory of the uncanny valley, then, the term "realism" needs to be better defined. Here we define realism as "being within the possible, naturally-occurring appearance of real human beings". Realism then can be considered across several dimensions including static, dynamic, and contextually responsive (i.e. contingent or interactive). Within each dimension, there will be many sub-characteristics of realism, such as physical geometry, texture, and coloration, which will be constrained by human biology.

Some characteristics of a figure can change without changing the level of realism. For example, real humans may be beautiful, comforting, or unsettling to behold, depending on variances that are all biologically possible. Likewise characters in the same cartoon oeuvre can be a scary villain or cute protagonist depending on design qualities, like Ursula versus Arial in Disney's *the Little Mermaid* [Thomas and Johnston, 1995].

A humanoid figure may exhibit extreme realism in some characteristics while deviating from realism in others (for example, a realistic face with a cartoon body). Alternately, the characteristics of a figure may evenly deviate from realism (like a face and a body that are both slightly cartoonish). With so many ways to deviate from realism, and so many ways to modulate aesthetic, it would seem plausible that human reaction could vary at any given level of realism.

If human reaction is indeed variable at any given level realism, this implies that the aesthetic space is more densely populated, more like a cloud of aesthetic possibilities rather than the definite curve drawn in Mori's uncanny valley graph. In [Hanson et al, 2005], we show that anecdotal examples seem to indicate that there can be indeterminately many possibilities for aesthetic humanlike depictions that do neatly populate Mori's curve of a valley (figure 1-c).

If all possible aesthetic permutations were included, then presumably the space would be filled densely. The curves in these figures would just represent select paths through the space of all aesthetic possibility. These curves, including the uncanny valley, would be merely trails of biased sample selections. The anecdotal examples imply that human aesthetic reaction is much more complex than Mori's uncanny valley would indicate, and not strictly related to realism.

Extending this concept, let us imagine a space of all possible aesthetic robot designs, plotted (like Mori's graph) against the two dimensions of "realism" and "reaction". This space of all possible aesthetic robot designs would include all possible permutations of numerous aesthetic variables (e.g. facial proportion, skin texture, body style, etc), some of which might modify a robot's appearance and movement without altering the level of realism, but which could dramatically shift people's reaction up or down. We consider that lower region of the aesthetic space would be uncanny, while the upper region of the aesthetic space would be appealing, consistent with Mori's graph. Assuming that aesthetics which are appealing (and not eerie) will facilitate better social interaction, we designate the upper region of the space the "zone of engagement".

The pivotal question regarding the truth of the uncanny valley, thus can be rephrased as whether the zone of engagement is continuous over the entire dimension of realism—from completely unrealistic to perfectly realistic. If one can demonstrate a continuum of attractive, low-eerie robots from highly abstract to fully realistic, then it will be shown that there is no valley. That is: there is no unavoidable dip into the eerie). The valley would be proven to be an illusion caused by biased sample selection.

In formulating our hypotheses regarding whether or not level of realism is the primary factor in the attractiveness/eerieness of humanlike figures, the science of human social perception provides some guidance regarding the multiple dimensions of attraction and repulsion.

2.1. Neural Templates for Attraction, Repulsion, and the Uncanny

Neuroscience supports the contention that people's reaction to appearance will be substantially related to factors other than realism—factors much more dimensionally complex than the theory of the uncanny valley implies.

While studies indicate that we are much more sensitive about real human faces [Gauthier et al, 2000], the specific forms of beauty and ugliness inspire remarkably consistent human responses, regardless of level of realism [Etcoff, 2000; Zaidel, 1997; Thomas and Johnston, 1995; Tong et al, 2000] scientific literature on facial attractiveness, shows that even among real humans, minor deviations in appearance can change a face from beautiful to ugly or disturbing [Etcoff, 2001; Cunningham et al, 2002]. These forms of beauty and revulsion are important to consider when designing socially engaging robots, and also must be considered as possible confounds when testing for human reaction to realism. A poorly designed humanlike figure can repel people, regardless of its level of realism. Good design is highly attractive.

Universally, clear skin, nice hair and large expressive features are considered attractive [Etcoff, 2000; Cunningham et al, 2002]. Likewise, the large eyes and forehead, and small nose and jaw associated with neoteny (the "baby scheme") are considered universally endearing and inspiring of protection [Eibl-Eibesfeldt, 1970; Etcoff, 2000; Cunningham et al, 2002; Breazeal, 2002]. In general, averaged faces are more attractive than the median (presumably by canceling unhealthy deviations from the norm) [Rubenstein et al, 2002; Rhodes et al, 2002]. However, average faces are not the most attractive. The most attractive faces deviate from the average but only in very specific ways, usually in features associated with neoteny, sexual maturity, or senescence [Cunningham et al, 2002; Etcoff, 2000]. Each of these exaggerated feature-sets inspires different behavior in humans. Neoteny features inspire nurturing, sexual maturity features inspire both sexual attraction and friendship, while senescence features inspire mentoring relationships [Cunningham et al, 2002; Zebrowitz and Rhodes 2002].

The beauty of a countenance signifies a history of genetic and environmental healthiness, in terms that would have been significant to our ancestors. We evolved to perceive strong upper bodies as attractive in men because this signifies capacity to fistfight, which in turn signifies likely social dominance (Etcoff equates pectoral muscles of men to the antlers of stags) [Etcoff, 2002]. Meanwhile, we evolved to perceive certain curves as attractive in women, because they signaled a strong likelihood of fertility [Etcoff, 2000; Grammer et al, 2002]. While most such signifiers may hold little meaning in our modern world, their recognition would have conferred benefits to our ancestors' survival, so would have been selected to become powerful, attention-grabbing aesthetic templates in our central nervous systems.

These templates of beauty persist as vestigial, eye-popping urges that will continue to shape our tastes no matter how hard we try to rationally over-ride them. Our preferences for beauty are, quite deeply, stamped in our brains. This is convenient for our purposes, as these preferences help to define the constraints of design of animated characters in robotics.

It is well demonstrated that human aesthetic preferences transfer to nonhuman objects and beings [Norman, 1992; Kanwisher, 1997; Breazeal, 2002; Fong et al, 2003]. People are more kind and protective of other species that are baby-like in appearance [Lorenz, 1946; Eibl-Eibesfeldt, 1970; Etcoff, 2000; Cunningham et al, 2002]. They purchase more clothes if worn by beautiful mannequins [Etcoff, 2000]. They more readily warm up to cartoons that are cute or beautiful [Thomas and Johnston, 1995]. We expect that these universal neural-templates of beauty may make robots more appealing too, regardless of level of realism.

Conversely, other aesthetic patterns are universally regarded as ugly, disturbing, or eerie. Sickly eyes, bad skin, extreme asymmetry, and poor grooming are all repulsive to people [Etcoff, 2000]. Generally, signs of illness or injury are found disturbing [Darwin and Ekman, 1872/1998; Etcoff, 2000]. Facial forms akin to expressions of terror, psychosis, and subterfuge are also found universally alarming [Darwin and Ekman, 1872/1998; Ekman, 1970; Adolphs et al, 1998]. Such eerie signifiers are used in cartoons and art to depict villains or monsters, and this trick works regardless of how abstract the cartoon.

Such negative features would certainly be associated with a "walking corpse"—Mori's example at the bottom of the purported uncanny valley. But as discussed, these features are not attached to a given level of realism, any more than a big smile is, or large cute eyes are, etc. Avoiding perceptual templates that trigger fear may help avoid the uncanny reaction, regardless of the level of realism.

The many dimensions of human attraction and repulsion can interact in complex ways at times—such as when percepts of attraction and repulsion are activated simultaneously, resulting in a figure *both* attractive *and* eerie (e.g. Tim Burton's *the Corpse Bride*).

But what about sensitivity to realism—is there any evidence that realism does make a difference?

Many studies show that people do show special sensitivity to the real human face [Tzourio-Mazoyer, 2002; Kanwisher, 1998;Kanwisher 2000;]. We are much more sensitive to familiar faces and objects [Gauthier, 1998], can better read faces with which we are familiar [Golby, 2001]. People appear to find more familiar types of faces to be more attractive [Reiman et al, 2000; Etcoff, 2000; Cunningham et al, 2002]. People are especially sensitive to subtleties of real human faces—moving one facial feature by just 1mm will transform a real face from attractive to unattractive [Etcoff, 2000]. These sensitivities imply that more realistic faces may trigger more demanding expectations for anthropomorphic depictions [Hanson et al, 2005]. Our discriminating taste for realism may mean that, very subtle aesthetic violations, which may go unnoticed in more abstract depictions, may be upsetting with a realistic depiction. In essence, realism may be fundamentally more challenging. But such sensitivity to realism may actually be useful—realism may be more communicative, conveying social data at higher bandwidths. Thus, increased sensitivity doesn't imply a valley, just more demanding design requirements as realism increases, but with possibly greater rewards.

There are other neural mechanisms that may generate something like the legendary uncanny valley effect.

Social recognition is processed in two distinct neural activities—one being low-resolution, motion-based face/gesture recognition (processed by the pSTS region of the brain), and the other being high-resolution, static recognition processed by the Occipital Cortex and in particular the Inferior Temporal (IT) lobes [Riesenhuber and Poggio, 2000]).

Processing static images is achieved the Inferio-Temporal lobe by close collections of cells that are tuned to encode various views of an object or figure [Riesenhuber and Poggio, 2000]. The static aspects of faces, figures and objects are processed by separate brain structures and distributed patterns of brain structures.

In this system, one brain organ—the lateral fusiform gyrus— is well-demonstrated to respond to invariant aspects of the human face (when contrasted with object or house recognition), and so has been dubbed the Fusiform Face Area (FFA) [Puce et al. 1995; Kanwisher et al. 1997; Kanwisher 2000]. Intimately interconnected with brain regions that manage lexical information, such as names and other language-related data [O'Toole et al, 2002], the location of the FFA (as well as the entire IT "what" system) allows fast matching between the expertise of invariant visual attributes and expert language relating to the figure being viewed [Lamb, 1999; Adolphs, 2001].

Other regions, notably the Occipito-Cortex zones M100, M170 [Kalanit Grill-Spector 2002; Liu et al, 2000], and the Occipital Face area (OFA) [Gauthier 2000, Helpert,1998, Haxby 1999] have been shown to be responsive to the human face, but without the capacity for expert identification that is exhibited by the FFA. Given that facial recognition may be useful to many social cognitive tasks, such redundant facial responsivity may offer improved localized computational efficiencies.

The FFA also appears instrumental in processing objects with which we are expert [Gauthier et al, 2000; Gauthier et al, 1998], in performing social attribution tasks [Schultz et al, 2003], and attribution of agency [Schultz et al, 2003].

Dynamic actions of the face and figure (including gaze-tracking, head-orientation, identification of facial expression movement and speech-related motions) are processed by the posterior superior temporal sulcus (pSTS) [Critchley et al. 2000; Hoffman & Haxby 2000; O'Toole et al. 2002]. While the pSTS, is uniquely responsive to motion, emotional expressions are also processed (seemingly as static percepts) by the amygdala, the anterior cingulated sulcus, the FFA, and language-related regions of the brain. Perception of face figure and object is not performed in isolated regions though. All of these organs operate in concert to achieve visual social cognition by coordinated, cascading actions among many structures [Castelli et al 2000]. [Haxby et al, 2001] found that images of faces, cats, houses, and objects (and noise) elicit distinct distributed patterns of neural firing, suggesting that such subjects are at least partially encoded in a distributed way. It may be that such distributed patterns encode semantic interrelationships among the diverse perceptual data of faces, places, and objects, and the diverse informational aspects of each (for example: a face's moving and invariant information [O'Toole, et al, 2002]).

Haxby's work provided the basis for [LaBar et al, 2003], which found that visual stimuli of moving expressions (fear, anger), activate distinct distributed patterns, involving varied structures that generally included the pSTS, the amygdala, and, most notably, the FFA. In this study, the FFA responded with greater activity to dynamic expressions than to static expressions, and this response was particularly punctuated when subject viewed fear expressions. [Kesler-West, 2001] also detected elevated FFA response to various social affects and facial expressions, including happy, sad, fearful, and angry expressions. As with LaBar, the FFA was particularly sensitive to fearful expressions, and as with LaBar, the amygdala also elevated in response to fearful expressions. It is notable that neither the amygdala nor the FFA responded especially to happy expressions. Happy expressions instead activated the medial frontal/cingulate cortex, an area found to be important in the initiation of language.

These findings of Kesler-West and LaBar imply that very different neural systems will be engaged by percepts of fearful expressions than by those engaged by happy or sad expressions. More generally, these findings may indicate that visual percepts that indicate social ease will trigger preparations for social engagement, while visual percepts associated with crisis will preempt social attention to prepare for danger.

[LaBar et al, 2003], also found that similar fear activations occur when subjects viewed a sliding identity morph, which the authors suggest may be a reaction to implausibility or to signals interpretable as subterfuge, innately associated with crisis via evolution. It does seem possible, however, that this response is caused by simple logical inconsistency, in essence causing a double-take, a rapid error-check. In both cases of crisis reaction, the FFA played a prominent role.

It is widely accepted that the amygdala plays a strong role in face-related social cognition [Schultz et al, 2003]. The amygdala also has been shown to be prominent in the distributed networks reported in [Kesler-West et al, 2001; LaBar et al, 2003]. [Schultz et al, 2003] further describes a distributed social network composed of the amygdala, medial prefrontal cortices, and the Fusiform Gyrae (FG) (including the FFA), and when changeable and dynamic aspects are involved, the pSTS. [Schultz et al, 2003], also found that the amygdala responds to neutral, non-emotional faces, for the first time implicating the amygdala in non-emotional face processing. Because the ventral cortices are highly integrated with the amygdala [Amaral & Price 1984] data may be exchanged at a high rate between the fusiform gyrae and the amygdala. The amygdala also reliably activates when subjects are asked to judge personality characteristics from images of faces [Adolphs et al. 1998; Baron-Cohen et al. 1999; Winston et al. 2002, Schultz et al., 2003]. The amygdala and its gateway, the anterior cingulate cortex, are well-known to produce the social feelings that are critical to learning and managing social complexities [D'Amasio, 1994; Adolphs et al., 1998].

[Schultz et al., 2003] found that the FFA operates in close tandem with the right amygdala in a distributed network to process social tasks. In tasks of social cognition, Schultz felt that the amygdala guides the FFA towards more fine-tuned, situational face-processing. In this light, the findings of LaBar and Kesler-West would indicate that this amygdala-FG feedback elevates with fear-associative face-image stimuli, causing the FG to search more closely for expert signs of danger and safety.

The brain generally processes objects, places and faces in very different regions than it does faces (Adolphs, 2001; Spiridon and Kanwisher, 2002). A humanlike robot that looks too mechanistic might trigger competition between the face and object regions, causing cognitive dissonance that might be disorienting or disturbing, an effect similar to that of LaBar's sliding identity morphs [LaBar et al, 2003]. A humanlike robot that straddles the categories of object and human might be upsetting. Our sensitivity to realism may render inhuman perceptual signatures more disturbing the more realistic the figure becomes. However, such a disturbing effect may not be related to *level* of realism, per se, but rather to an awkward mixing of conflicting perceptual cues. If so, then to avoid the effect, one would designing a robot without the problematic mechanistic artifacts.

In summary, the human brain has a social-regulating emergency alarm—a binary gate that enables social engagement under the right perceptual conditions, but will divert attention toward fearful, self-protection under the wrong perceptual conditions.

Is some level of near-realism, in and of itself, such a wrong perceptual condition? That is, is there an inherent “uncanny valley”?

To test this question, we posit several hypotheses regarding the uncanny valley, build robots that attempt to challenge the uncanny valley, and then test these robots with human subjects.

We hypothesize first that human reaction is at least partially unrelated to level of realism. Second, we hypothesize that at least one continuum of positive reaction exists across the dimension of realism, thus disproving the valley. Third, we hypothesize that the aesthetic reaction space is densely populated and not neatly condensed into a curve, Mori’s or otherwise. Fourth, we hypothesize that the human reaction to anthropomorphic depictions will be affected by three main factors: human neural templates for beauty, neural templates for signs of danger (including signs of illness or morbidity), and by novelty/acclimation effects.

To test these hypotheses, we have needed robots that go into the territory proscribed by the uncanny valley.

2.2. Sending Robots in to Explore the Valley

Historically, the entertainment industry has most aggressively explored realistic and nearly realistic robotic hardware, for use in movies and theme parks. Such entertainment devices are known as “*animatronics*”. These machines have taken a wide diversity of form, from the realistic “Abe” Lincoln of Disneyland, to the bizarre aliens of the *Men in Black* movies.

Such hardware has not generally been used with social robotics research, with two notable exceptions being Cynthia Breazeal’s collaboration with Stan Winston on the Leonardo robot, and Hiroshi Ishiguro’s work with Kokoro Dreams on robots including the Repliee Q1 (see figure 2).

Figure 2. Hiroshi Ishiguro and the Repliee robot, 2005.

Animatronics tend to be expensive to design and build (restricting research and creativity), limited in expressivity (the faces don’t make a full range of humanlike expressions), and power hungry (limiting mobility and fusion with expressive humanoid bodies like the Asimo or Qrio). Altogether, the limitations of animatronics technology prevent expansive use in social robotics, and generally limit usefulness outside niche applications of films and theme parks.

Animatronics’ problems of costliness, low expressivity, and power consumption all result from the physical dissimilarity of animatronics’ simulated facial soft tissues relative to real human tissues. Human facial soft tissues are mostly liquid, filling cellular membranes like billions of tiny water balloons. The liquid molecules will slide into any geometry that the membranes can tolerate. In this way, the human face is like a sealed wet sponge. Animatronic facial tissue on the other hand, is made of solid elastomer (e.g. rubber), which is composed of tangled, spring-like polymer molecules that unwind when elongated but are geometrically interlocked. Thus, these molecules are fundamentally restricted from reproducing the geometric plasticity of human facial tissues. In effect, the force required to expressively move animatronics materials are orders of magnitude above that required by facial soft tissues [Pioggia et al, 2003].

To resolve these issues, the author has developed a series of methods for making sponge-like elastomer materials that move more like facial soft-tissues. These materials, which we call "Frubber" (a contraction of *flesh* and *rubber*), wrinkle, crease, and amass, much more like skin than do animatronics materials (see figure 3). They also consume very little power—less than 10W when affecting a full range of facial expressions and speech-related mouth motions. In our tests, the material requires less than 1/22nd the force and energy to move into facial expressions relative to animatronics materials.

Figure 3. Frubber robot expressions, consuming less than 10 watts.

The reduced energy consumption enables use with battery-powered biped walking. Being a porous, Frubber also weighs much less—also a benefit for untethered walking robots. Such integration with a walking gestural body will be significant as it allows the exploration of aesthetic of the entire, integrated humanoid figure as an autonomous social being. We may expect a gestural walking body to affect how people react to a social robot. Moreover, bipedal walking will certainly impact the performance capability of robots in our world at large, expanding the range of possible social engagement and interaction between robots and people.

Since spring 2002, the author has realized ten iteratively more expressive robot faces using Frubber materials. In the process the author has also pushed the rapid design and manufacturability of the robots, seeking to bring the robots to mass production and low cost [Hanson , 2004; Hanson et al 2005]. This is significant because to unabashedly explore the design potential of robots, the technology must be fast, easy, and cheap to customize and use. Additionally, the social AI software must be refined to facilitate improved social engagement. We expect that a less expensive, yet more design-flexible, hardware platform will generally enable more widespread software development for this kind of robot. To this end, we are developing a host of rapid design, animation, and AI development tools to produce new robots quickly and affordably.

To make the robots more engaging and seemingly intelligent, the author's team has implemented AI-driven intelligent interactive software using face tracking, face recognition, automatic speech recognition, natural language processing, and speech synthesis. Much of this software is available open-source at the author's website, www.hansonrobotics.com.

In the author's android portrait of sci-fi writer Philip K Dick (PKD-A) (see figure 4), the PKD-A software (programmed by Andrew Olney of the University of Memphis) also incorporates thousands of pages of the author's writing into a conversational sketch of the deceased sci-fi writer's mind. Such literary and narrative aesthetic can be the key to designing humanlike robots that are socially engaging—a good personality goes a long way. This technology combination in PKD won the AAAI 2005 first place award for Open Interaction.

Figure 4. Philip K Dick Android talking with a child.

In June 2005, the author combined robot Eva with the walking biped robot Hubo of KAIST, producing what may be the world's first expressive human face on a walking biped robot. It is the lightweight, low-power benefits of the Frubber material that enable such performance.

The author's latest robot, a fully expressive portrait of Albert Einstein, further demonstrates the significance of Frubber, mounting atop another Hubo that is customized to the character "Albert-Hubo" (see figure 5). Unveiled at the APEC summit Nov. 2005 in Busan, Korea, the Albert Hubo greeted world leaders, schoolchildren, the public at large, and the press media over the span of 5 days of presentations. The head of the robot, designed and built by the author's company Hanson Robotics Inc to be as realistic as possible, demonstrates realism in static expressions, dynamic action, and (to some extent) interaction.

Figure 5. Albert Hubo.

The aesthetic of the body of the Albert Hubo, was designed by KAIST in collaboration with Won Seup Kim to be aesthetically beautiful, but is *not* human in numerous regards: the hips are absent, the feet are strange and square, and the joints are mechanical, and the hands are bulky in an inhuman way.

When combined with the realistic head, the integrated robot should be down deep in the uncanny valley. And yet the robot was extremely popular and engaging. Indeed, resoundingly, he was the hit of the show at the APEC summit [U.S. News, CNN, and MSNBC, Nov. 2005].

Hanson robots attempt to prove that nearly-realistic robots can be engaging.

Encountering these robots, people generally state that they find the robots as “cute” and even “lovable”. Generally people have not regarded these robots as horrific or akin to walking corpses. We seek to explore robotic aesthetics both informally—by building robots—and formally—by putting the actual robots into human subject tests.

The human subject tests described in this paper utilize the Philip K Dick android (PKD-A).At this time, the Albert Hubo is also available to researchers for cognitive science and/or software development (please enquire with the author with special requests). Custom robots are also available.

2.3. Early tests to Explore the Valley

In November 2004, we conducted two web-surveys that sought to test the uncanny valley premise. The first showed videos of two Hanson robots, animated to simulate humanlike facial expressions, and asked what people thought of the images (see figure 6). The second survey showed a continuum of humanoid depictions, shifting from cartoonish to realistic over six frames (fig. 14).

The robots in the first survey consisted of “Eva”, an adult female robot (figure 6 a), and “Yargh”, an androgynous pirate with rough skin (see figure 6 b). In some ways, both faces are very realistic both in static and dynamic appearance. Yet in other regards, the robots were very unrealistic: the pirate lacked the back and top of his head, Eva's eyes are strange and inhuman. Both lacked bodies and shoulders. These deviations are significant in that they would be fatal if present in a human, so may connote mortality to a viewer.

Reactions to each robot were similar: more than 80% of respondents found the stimuli "entertaining", 73% found the stimuli "appealing", and over 85% found the robots to look "lively" and "not dead". While no respondents reported that humanlike robots are disturbing, 81% said that they found the video "not disturbing", with almost 29% reported that they found the videos "at least somewhat disturbing" [Hanson et al 2005]. This was in spite of numerous unrealistic features (head on a stick, no back of head, etc).

Figure 6. a. video 1, Eva emoting;: **b.** video 2, robotic pirate gesturing.

In the second test, we showed human subjects a morph from a cartoon human (Disney's Jasmine) to a real human face of actress Jennifer "Love" Hewitt (see figure 7).

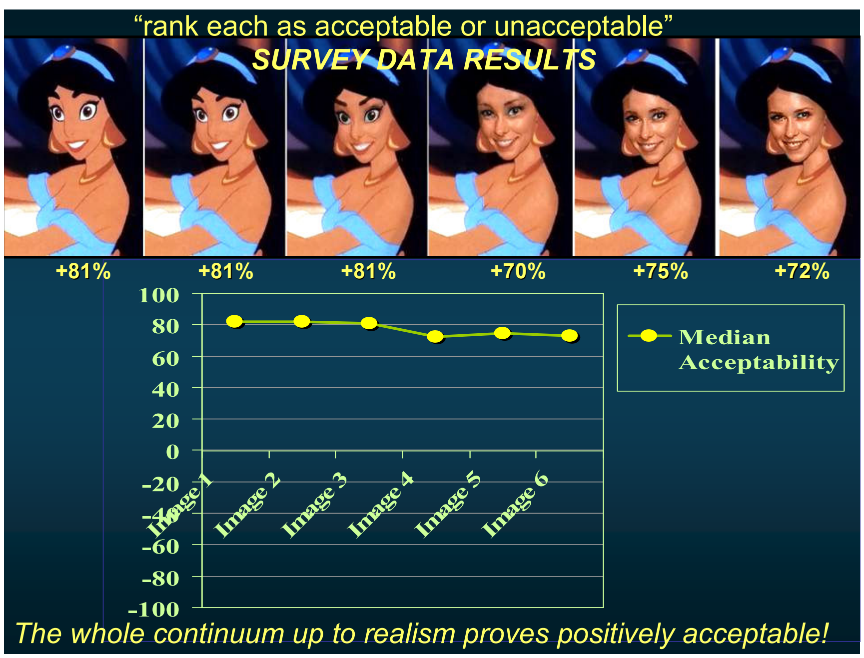

Figure 7: Results of test 2.

The results of the second survey clearly show that viewers found the whole continuum to be positively acceptable. The reaction never dipped into the negative region, thus showed no sign of the repulsion that defined the "valley" of Mori's uncanny valley.

In these studies, neither dynamic nor static images appeared to result in an inherent valley effect. However, the images of the robots do not represent the continuum needed to test the valley premise, and the static images are not robots. To test real robots in morphs, we recently ran a new series of experiments.

3. Experiments and Results

To further test the uncanny valley theory, in October of 2005 we administered a new series of assays wherein we showed human subjects series of images of robots and humanlike images, with varying levels of realism, and varying aesthetic qualities.

The test consisted of a series of images that morph from abstract robots, to our realistic robots, to images of real humans. The human subjects were asked to rank the images from 1 to 10, on several metrics: realism, appeal, eeriness, and familiarity.

In these experiments our control morph (see figure 8) was inspired by a morph used in the experimental work of Hiroshi Ishiguro and Karl MacDorman, which contains a continuum of morphed images that elicit reactions from subjects which follow the pattern predicted by the uncanny valley.

In our experimental morph, meanwhile, the morph images were designed with the intention to make them appealing and not eerie (see figure 9). If human subjects reacted with consistently low-eerieness ratings, then this would imply that the uncanny valley is avoidable, at least in the static domain.

In a third morph, we attempted to show that other examples would invert the uncanny valley, so that the abstract robot and the real human are the most eerie, and that the nearly humanlike robot is the least eerie (figure 10).

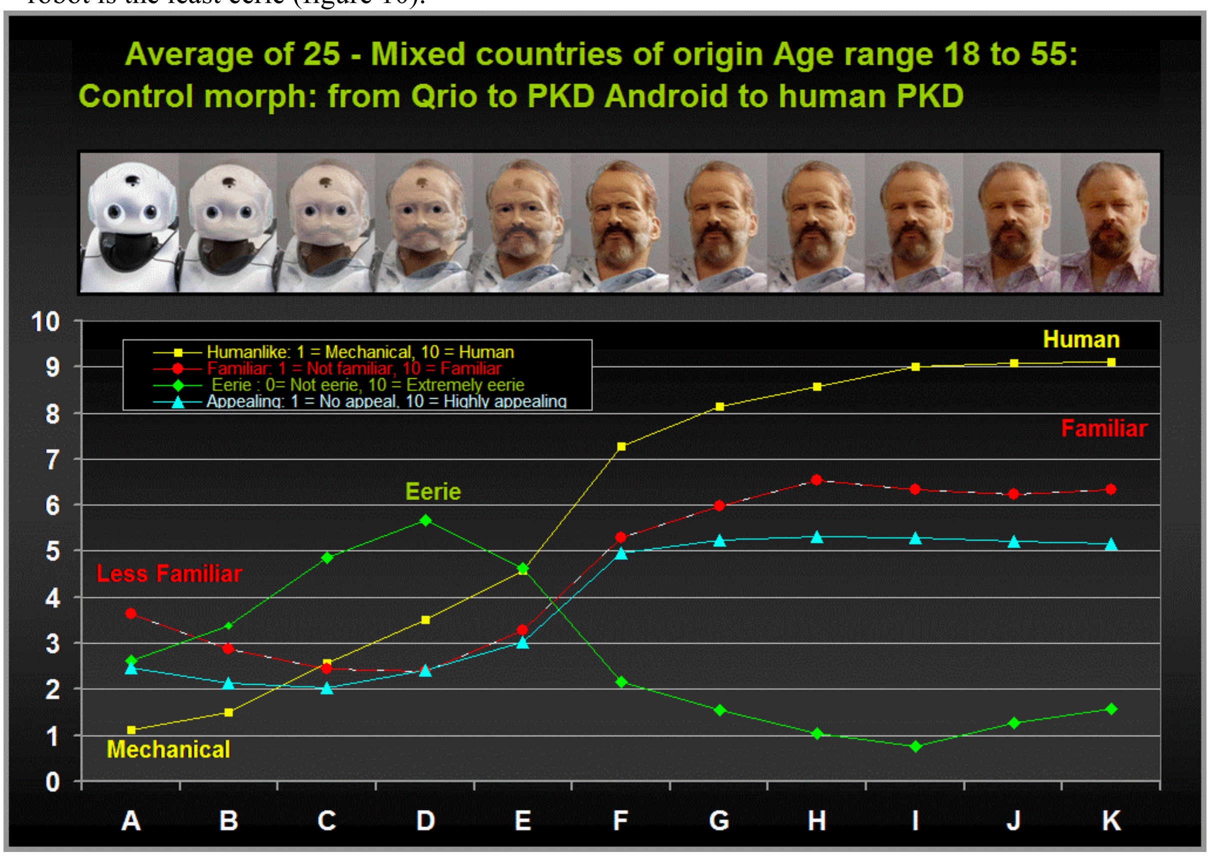

Figure 8: A morph that returns an apparent uncanny valley effect.

We showed the images to 25 test subjects. Reaction to the control figures followed the pattern predicted by the uncanny valley theory.

Reactions to the tuned morph, however, were striking in that the attractively-tuned figures were found to be consistently low in eeriness and high in appeal.

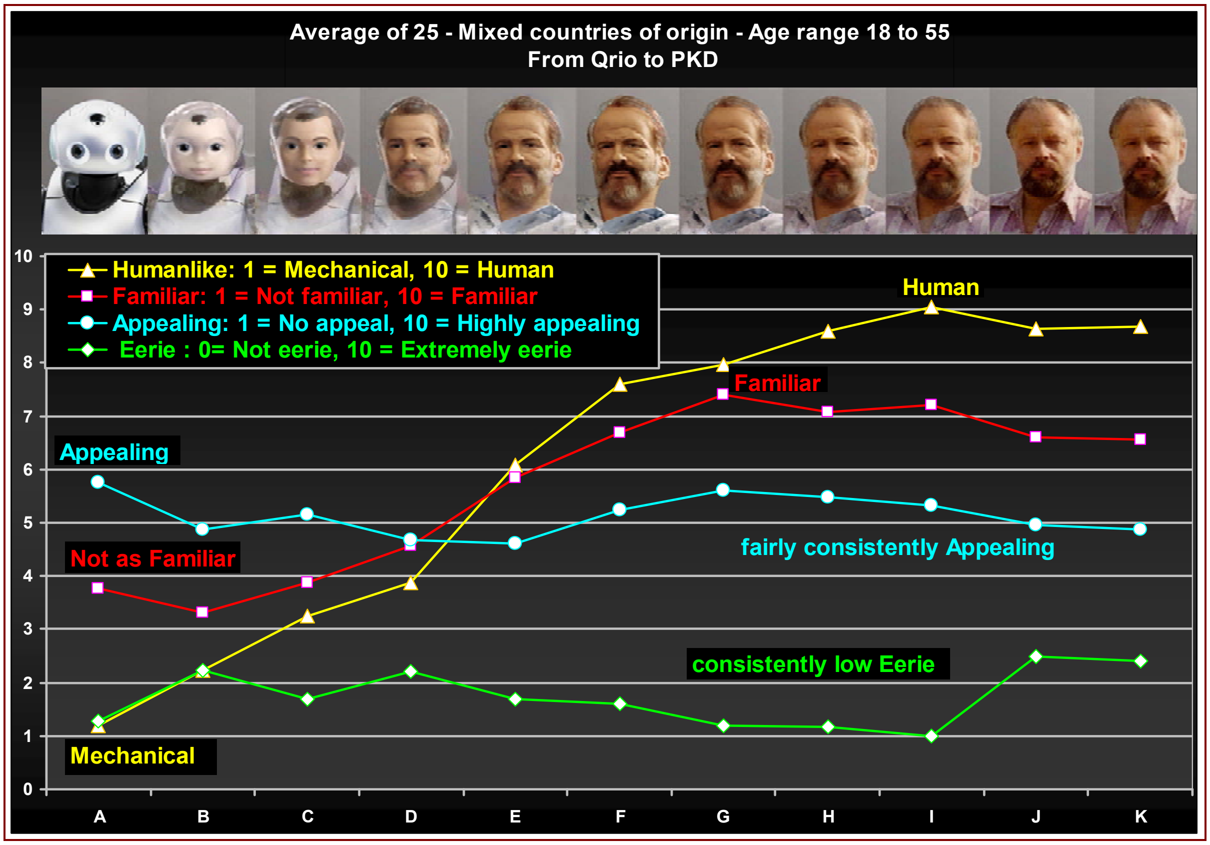

Figure 9: A morph that returns little or no uncanny valley effect.

This strongly implies that reaction is at least partially decoupled from realism. These results also imply that, with well-tuned faces, there can exist a continuum of appealing anthropomorphic depiction across the range of realism, thus supporting the contention that there is no inherent uncanny valley.

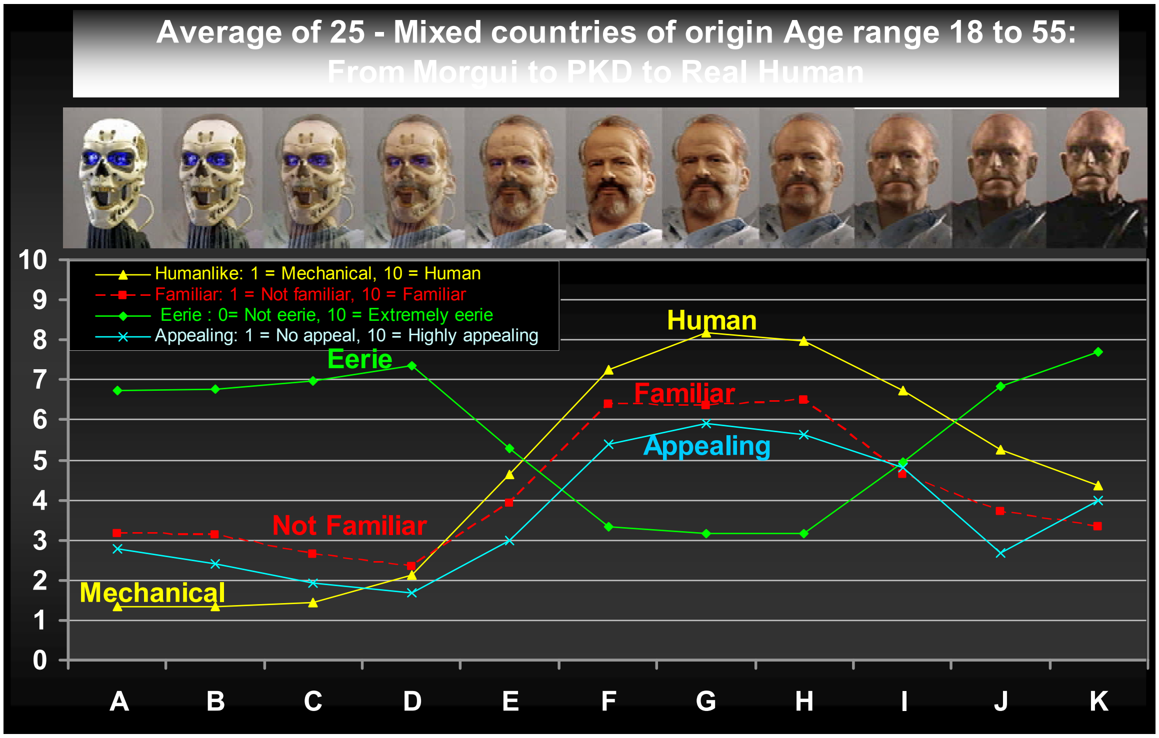

Figure 10: A morph that inverts the uncanny valley effect.

When viewing the third morph, people found the very abstract robots and real humans to appear very eerie, but found the nearly-human looking robot in the middle to be less eerie by contrast. This follows a pattern that is the reverse of that predicted by the uncanny valley theory.

However, these results are consistent with the contention that certain characteristics of average human appearance will be found attractive regardless of the level of realism, and that deviation from this average will be unappealing unless aesthetically well-tuned.

The examples chosen are not animated, and so we do not know if the results will be very different for animated figures, so further testing is required. However, as a preliminary body of evidence, it seems strongly that the uncanny valley may not be true.

The uncanny valley theory is not inconsistent with the data results. In anticipation of further evaluations, we are proposing a new theory.

4. In pursuit of a better theory of robot design

We take the existing neuroscience of facial perception as the groundwork for our alternative theory,

Neuroscience demonstrates existing neural-templates for beauty, ill health, and for “things we are or are not accustomed to”. While people are responsive to cartoons, they are much more sensitive to the real human face [ref], especially facial types with which they are familiar. So, the more realistic a face, the more a person may expect of the robot. This may make realism challenging, but it may make realism communicative at a higher bandwidths.

More realistic faces on robots may result in human-identity challenging effects, such as implications of subterfuge. People use different brain regions to process objects and people. However, the FFA processes both faces and social attribution (“agency”). When an object (such as a robot) too closely resembles a person, we may expect a sensation of category-mismatch between perceptions of “object” and “person”, perhaps causing the disorientation and potential alarm that we call “uncanny”.

Such effects are not well-defined or parameterized, and warrant further investigation. Intuitively, it seems that objects that too closely pose as humans may challenge our sense of human identity. They may be read as imposters. If they break or show signs of being a mechanism, they may activate latent fears of mortality. The only way to test these effects is to build more robots that should be in the uncanny valley, but design them extremely well, and see if people find them eerie.

If the illusion of life can be created and maintained, it seems possible at least that the uncanny effects can be mitigated. It may be that any level of realism or can be socially engaging if one designs the aesthetic well. This, in effect, would represent a bridge of good aesthetic, which inspires us to name the revised theory the "Path of Engagement" (POE).

This POE theory has some interesting ramifications.

If people are indeed more sensitive to realistic faces, this may imply that realistic faces transmit a rich, high-bandwidth stream of data. Conversation diverts attention from watching for danger, so a face that behaves strangely or in an unhealthy manner may trigger survival or fear reflexes. Alternately, it may trigger "surreal" (dreamlike) feelings, rather than fear. Thus, people may find the robot strange but not frightening. As no "valley" is inherent; anthropomorphic depictions can be either disturbing or appealing at every level of abstraction or realism. People simply get more sensitive with increasing levels of realism.

In future work it will be important to push the science of the art (exploring the neuro-cognitive basis of social aesthetics), and the art of the science (art as a black-box problem solver for problems of social intelligence). Certainly, the science, art and technology of social robots will benefit from further exploring the boundaries uncanny valley. If the valley proves false, then the world of robot design can only benefit from the loss of an artificial proscription.

5. Conclusion and Future Work

As robots proliferate, they will more frequently engage people in face-to-face interactions. The success of such encounters will depend substantially upon the aesthetics of a given robot.

Identification of fundamental principles of robot aesthetics can greatly accelerate the successful deployment of robots. On the other hand, a specious principle may artificially constrain the design options and decrease the use of robots in the world.

If robots get to the point where they seem alive in every sense of the word, and are well designed to be friendly and cute, then people won't feel like the robots are half-dead. At that point even nearly realistic robots won't be so eerie.

Presently even the most realistic robots may seem partly-dead, because in many ways they are. They are only partly aware. They shut down instead of going to sleep, and then they sit there frozen. They break. These flaws combined with a humanlike appearance, can remind us of our own mortality. They also may seem like non-human matter impersonating humans, conveying the threat of an imposter.

Removing these flaws and making them alive, friendly, and attractive, and I suspect the level of realism will not matter. We will quickly feel more comfortable with them.

About half of those informally surveyed after interacting with the Hanson robots reported no fear reaction to the robots. Almost all seemed to enjoy the experience of encountering the robots. It would seem that the uncanny valley effect won't get in the way of nearly human robots being desirable in our lives.

There may always be people who are disturbed by robots that look too humanlike. We may want to consider that we can't please all the people all the time.

Ultimately, good design can help to make robots lovable and enter the human family. More freely exploring the full range of robot aesthetics will certainly accelerate the evolution of humanoid robot design. Moreover, the expanded exploration promises to help us better understand human social perception and cognition.

6. Acknowledgements

The author would like to acknowledge the help and guidance of Karl Macdorman, Alice Otoole, Thomas Linehan, and Dennis Kratz. The author especially thanks Elaine Hanson, the author's mother, for her design and layout of figures 8, 9, and 10, and her hard work in administering the surveys to human subjects.

7. References

Adolphs, R., Tranel, D., & Damasio, A. R., "The Human Amygdala in Social Judgment." Nature, 393, 470-474; 1998.

Adolphs, R. (2001). The neurobiology of social cognition. Current Opinion in Neurobiology, l1, 231-239.

Breazeal C., Designing Sociable Robots, MIT Press (2002).

Caporael, L. R. Anthropomorphism and mechanomorphism: Two faces of the human machine. Computers in Human Behavior, pp185-211. 1990

Castelli, F., Happe, F., Frith, U. & Frith, C., Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. NeuroImage 12, 314-325, 2000.

Crosson, B., Sadek, J. R., Bobholz, J. A., Gökçay, D., Mohr, C. M., Leonard, C. M., Maron, L., Auerbach, E. J., Browd3, S. R., Freeman, A. J., Briggs R.W., Activity in the Paracingulate and Cingulate Sulci during Word Generation: An fMRI Study of Functional Anatomy Cerebral Cortex, Vol. 9, No. 4, 307-316, June 1999

Cunningham, M.R., Barbee A.P., Philhower C., "Dimensions of Facial Physical Attractiveness: The Intersection of Biology and Culture" from Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.

Damasio, A.R., The Feeling of What Happens: Body and Emotion in the Making of Consciousness, Harcourt Brace, New York, 1999, 2000.

Darwin, C., Ekman, P. (Ed.), The Expression of the Emotions in Man and Animals, Oxford University Press, New York (1998/1872).

Dion K.K. "Cultural Perspectives on Facial Attractiveness" from Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.

Duffy, B.R., Anthropomorphism and The Social Robot www.medientage-muenchen.de/archiv/pdf_2002/Duffy_12.2.pdf

Eibl-Eibesfeldt I, "Ethology. The Biology of Behavior". Holt, Rinehart and Winston, Inc., New York, 1970.

Ekman and Friesen, Basic Emotions (Ekman & Friesen, 1971)

Enquist M., Ghirlanda S., Lundqvist D., and Wachtmeister C. "An Ethological Theory of Attractiveness"

Etcoff, N. Survival of the Prettiest, Anchor, 2000.

Fong, T., Nourbakhsh, I., Dautenhahn, K. .A survey of socially interactive robots. Robotics and Autonomous Systems 42, 143.166, 2003.

Gauthier, I., Skudlarski, P., Gore, J.C., & Anderson, A.W. "Expertise for cars and birds recruits brain areas involved in face recognition". Nature Neuroscience, 3(2): 191-197, 2000.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. "Training "Greeble" experts: A framework for studying expert object recognition processes". Vision Research, Special issue on "Models of Recognition", 38: 2401-2428, 1998.

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y. & Eberhardt, J. L. Differential responses in the fusiform region to same-race and other-race faces. Nature Neuroscience, 4, 845 - 850, (2001).

Grammer K., Fink B., Juette A., Ronzal G., and Thornhill R. "Female Faces and Bodies: N-Dimensional Feature Space and Attractiveness" from Facial Attractiveness: Evolutionary,

Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.
Hanson, D, Olney, A, Zielke, M, Pereira, A. "Upending the Uncanny Valley", AAAI conference proceedings, 2005
Hanson, D. Rus, D., Canvin, S., Schmierer, G., Biologically Inspired Robotic Applications, in Biologically Inspired Intelligent Robotics, SPIE Press, 2003
Hanson, D., "Identity Emulation Facial Expression Robots", proceedings of American Association for Artificial Intelligence,Conference, August, 2002.
Hanson D. and Pioggia G., "Entertainment Applications for Electrically Actuated Polymer Actuators," in Electrically Actuated Polymer Actuators as Artificial Muscles, SPIE PRESS, International Society of Optical Engineers, Washington, USA, Vol. PM98, Ch. 18, March 2001.
Hanson D., Pioggia G., Bar-Cohen Y., De Rossi D., "Androids: application of EAP as artificial muscles to entertainment industry," Proc. SPIE's Electroactive Polymer Actuators and Devices Conf., 7TH Smart Structures and Materials Symposium, Newport Beach, USA, 2001.
Hara, F., Kobayashi, H., Iida, F., Tabata, M. Personality characterization of animate face robot through interactive communication with Human 1st Int'l W/S in Humanoid and Human Friendly Robots, pg 1-10, 1998.
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, 2425-30.
Heider, F. & Simmel, M. 1944 An experimental study of apparent behavior. Am. J. Psychol. 57, 243-259.
Kanwisher, N., McDermott, J., & Chun, M. M. The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17, 4302-4311, 1997.
Keating C.F. "Charismatic Faces: Social Status Cues Put Face Appeal in Context" from Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.
Kesler-West, M.L., Andersen, A.H., Smith C.D., Avison, M.J., Davis, C.E., Kryscio, R.J., Blonder, L.X. (2001). Neural substrates of facial emotion processing using fMRI. Brain Res Cogn Brain Research, 11, 213-26.
Kurzweil, Ray The Age of the Spiritual Machines Viking Press, 1999.
LaBar, K. S., Crupain, M.J. Voyvodic, J.B., & McCarthy, G., "Dynamic Percpetion of Facial Affect and Identity in the Human Brain. Cerebral Cortex, 13, 1023-1033, 2003.
Lamb, S. M., Pathways of the Brain, The Neurocognitive Basis of Language, Amsterdam & Philadelphia: John Benjamins Publishing Co., (1999).
Levenson, R., Ekman, P., and Friesen, W. "Voluntary facial action generates emotion-specific autonomic nervous system activity." Psychophysiology, 27(4): 363-383. 1990.
Little A.C., Penton-Voak I.S., Burt D.M., and Perrett D.I. "Evolution and Individual Differences in the Perception of Attractiveness: How Cyclic Hormonal Changes and Self-Perceived Attractiveness Influence Female Preferences for Male Faces" from Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.
Liu, J., Higuchi, C. M, Marantz, A., Kanwisher, N., "The Selectivity of the Occipitotemporal M170 for Faces", Neuroreport. Feb 7;11(2):337-41; 2000.
Masschelein Anneleen, "A Homeless Concept, Shapes of the Uncanny in Twentieth-Century Theory and Culture", Image & Narrative, Issue 5, January, 2003
Menzel, P., D'Aluisio, F. Robo sapiens: Evolution of a New Species, Boston, MIT Press, (2000).
Mori, Masahiro (1970). Bukimi no tani [the uncanny valley]. Energy, 7, 33–35. (In Japanese).

Norman, D. Turn Signals Are the Facial Expressions of Automobiles, Perseus Publishing, Cambridge, MA. 1992.
Ochsner, K. N., & Lieberman, M. D. (2001) The emergence of social cognitive neuroscience. American Psychologist, 56, 717-734.
O'Toole, A.J., Roark, D. A., Abdi, H., "Recognizing moving faces: A psychological and neural synthesis", The University of Texas at Dallas, March 13, 2002
Puce, A., Allison, T., Gore, J. C. & McCarthy, G. 1995 Facesensitive regions in human extrastriate cortex studied by functional MRI. J. Neurophysiol. 74, 1192-1199.
Reichardt, J., Robots: Fact, Fiction, and Prediction, Penguin Books Ltd., Harmondsworth, Middlesex, England, 1978
Rhodes G, Zebrowitz L. A. Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.
Rubenstein A.J., Langlois J.H., Roggman L.A. "What Makes A Face Attractive and Why: The Role of Averageness in Defining Facial Beauty" from Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.
Rhodes G, Harwood K., Yoshikawa S., Nishitani M., McLean I., "The Attractiveness of Average Faces: Cross-Cultural Evidence and Possible Biological Basis" from Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.
Reiman, E.M., Lane, R.D., Ahern, G.L., Schwartz, G.E., Davidson, R.J., Friston, K.J., Hart, A.J., Whalen, P.J., Shin, L.M., McInerney, S.C., Fischer, H., & Rauch, S.L. Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. Neuroreport, 11, 2351-2355, 2000.
Riesenhuber, M., Poggio, T, "Models of object recognition", Nature America, 2000.
Schultz R.T., Grelotti D.J., Klin A, Kleinman J., Van der Gaag C., Marois R., Skudlarski P., "The role of the fusiform face area in social cognition", Phil. Trans. Royal Society Lond. 358, 415–427, pub. online 21 J. 2003.
Shelley, M.W., Frankenstein, 1831.
Spiridon, M. and Kanwisher, N. "How Distributed Is Visual Category Information in Human Occipito-Temporal Cortex? An fMRI Study", Neuron, Vol. 35, 1157-1165, September 12, 2002, Copyright .2002 by Cell Press.
Thomas, F., Johnston, O., The Illusion of Life: Disney Animation, Hyperion; ISBN: 0786860707; Revised edition, 1995.
Tomlinson, B., "Dead Technology." Style Vol. 33 No. 2, p. 316-335, 2000.
Tong F., Nakayama, K., Moscovitch, M., Weinrib, O., Kanwisher, N. Response Properties of the Human Fusiform Face Area, Cognitive Neuropsychology, 17
(1/2/3), 257-279, 2000.
Tzourio-Mazoyer, N, De Schonen, S, Crivello, F, Reutter, B, Aujard, Y, Mazoyer, B "Neural Correlates of Woman Face Processing by 2-Month-Old Infants" NeuroImage 15, 454-461 (2002)
Weschler, L., "Why is This Man Smiling? Digital animators are closing in on the complex systems that make a face come alive." Wired Magazine, Issue 10.06 - Jun 2002
Yamasaki, H., LaBar, K.S. & McCarthy, G. "Dissociable Prefrontal Brain Systems for Attention and Emotion". Proceedings of the National Academy of Sciences, USA, 99, 11447-11451, 2002.
Zaidel, D. W. "Art and science" Nature 390: 330-330; 1997.
Zebrowitz L.A., Rhodes G. "Nature Let a Hundred Flowers Bloom: The Multiple Ways and Wherefores of Attractiveness" from Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing, Westport, CT. ISBN: 1-56750-637-2, 2002.