Monkey visual behavior falls into the uncanny valley

Shawn A. Steckenfinger and Asif A. Ghazanfar[1]

Neuroscience Institute, Departments of Psychology and Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08540

Communicated by Charles G. Gross, Princeton University, Princeton, NJ, September 3, 2009 (received for review July 19, 2009)

Very realistic human-looking robots or computer avatars tend to elicit negative feelings in human observers. This phenomenon is known as the ''uncanny valley'' response. It is hypothesized that this uncanny feeling is because the realistic synthetic characters elicit the concept of ''human,'' but fail to live up to it. That is, this failure generates feelings of unease due to character traits falling outside the expected spectrum of everyday social experience. These unsettling emotions are thought to have an evolutionary origin, but tests of this hypothesis have not been forthcoming. To bridge this gap, we presented monkeys with unrealistic and realistic synthetic monkey faces, as well as real monkey faces, and measured whether they preferred looking at one type versus the others (using looking time as a measure of preference). To our surprise, monkey visual behavior fell into the uncanny valley: They looked longer at real faces and unrealistic synthetic faces than at realistic synthetic faces.

animacy | audiovisual speech | avatar | face processing | human robot interaction

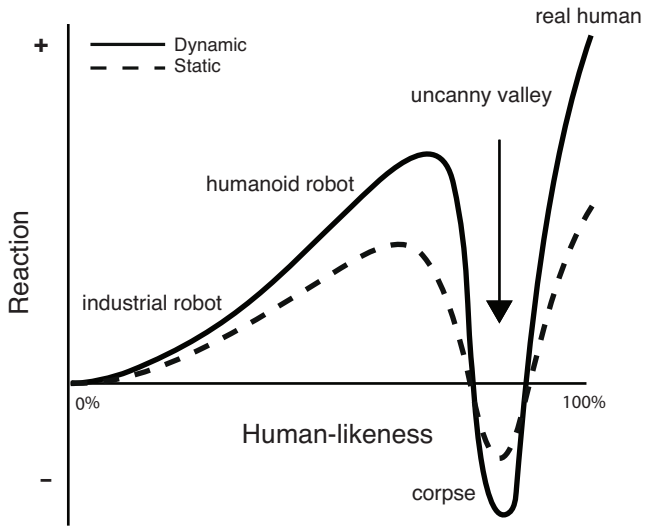

Fig. 1. A plot of hypothetical uncanny valley of perception/preference.

It is natural to assume that, as synthetic agents (e.g., androids or computer-animated characters) come closer to resembling humans, they will be more likely to elicit behavioral responses similar to those elicited by real humans. However, this intuition is only true up to a point. Increased realism does not necessarily lead to increased acceptance. If agents become too realistic, people find them emotionally unsettling. This feeling of eeriness is known as the ''uncanny valley'' effect and is symptomatic of entities that elicit the concept of a human, but do meet all of the requirements for being one. The label is derived from a hypothetical curve proposed by the roboticist, Mori (1), in which agents having a low resemblance to humans are judged as familiar with a positive emotional valence, but as agents' appearances become increasingly humanlike, the once positive familiarity and emotional valence falls precipitously, dropping into a basin of negative valence (Fig. 1). Although the effect can be elicited by still images, Mori further predicted that movement would intensify the uncanny valley effect (Fig. 1). Support for the uncanny valley has been reported anecdotally in the mass media for many years (e.g., reports of strong negative audience responses to computer-animated films such as *The Polar Express* and *The Final Fantasy: The Spirits Within*), but there is now good empirical support for the uncanny valley based on controlled perceptual experiments (2, 3).

Despite the widespread acknowledgment of the uncanny valley as a valid psychological phenomenon (4–7), there are no clear explanations for it. There are many hypotheses, and a good number of them invoke evolved mechanisms of perception (2, 6). For example, one explanation for the uncanny valley is that it is the outcome of a mechanism for pathogen avoidance. In this scenario, humans evolved a disgust response to diseased-looking humans (8), and the more human a synthetic agent looks, the stronger the aversion to perceived visual defects–defects that presumably indicate the increased likelihood of a communicable disease. Another idea posits that realistic synthetic agents engage our face processing mechanisms, but fail to meet our evolved standards for facial aesthetics. That is, features such as vitality, skin quality, and facial proportions that can enhance facial attractiveness (9, 10) may be deficient in synthetic agents that elicit the uncanny valley effect. Thus, a computer-animated avatar with pale skin might appear either anemic (and thus, unhealthy), unattractive, or both. Regardless of the specific underlying mechanism, in all cases of the uncanny valley effect, the appearance of the synthetic agent somehow falls outside the spectrum of our expectations–expectations built from everyday experiences with real human beings.

To date, there are no tests of these evolutionary hypotheses. The only way to test them directly would be to examine the behavior of a nonhuman species. Such an experiment would address the putative evolutionary origins behind the uncanny valley by examining whether it is based on human-specific mental structures. We investigated whether the visual behavior of a closely-related, nonhuman primate species (macaque monkeys, *Macaca fascicularis*, $n = 5$) would exhibit the uncanny valley effect. Like human studies (2, 3), we compared monkey preferences to different render-types: real monkey faces, realistic-looking synthetic agent faces, and unrealistic synthetic agent faces. The facial expressions for all three render-types included a ''coo'' face, a ''scream'' face, and a ''neutral'' face (Fig. 2*A*; neutral face not depicted). These faces were presented in both static and dynamic forms. Just as humans tend to look longer at attractive versus unattractive faces (11), monkey preferences were measured using both the number of fixations made on each face as well as the duration of fixation (12). If monkey visual behavior falls into the uncanny valley, then they should prefer to look at the unrealistic synthetic faces (because it does not appear to be a real conspecific) and real faces more than to the realistic

Author contributions: A.A.G. designed research; S.A.S. performed research; A.A.G. contributed new reagents/analytic tools; S.A.S. analyzed data; and S.A.S. and A.A.G. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: asifg@princeton.edu.

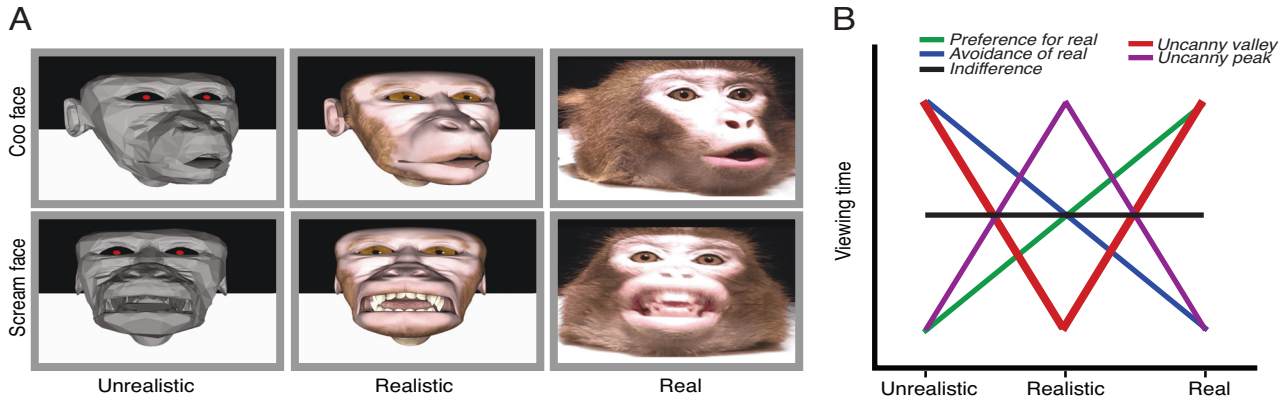

Fig. 2. Render-types and hypothetical outcomes. (*A*) The facial expressions and render-types used in the experiment. These images were extracted from the dynamic videos and represent the point of maximal expression. These frames were also used as the static stimuli. (*B*) Five hypothetical outcomes for the monkeys' visual behavioral response.

synthetic agent face (which elicits the concept of a real conspecific, but fails to live up to the expectation).

Results and Discussion

Five behavioral outcomes were possible in this experiment (Fig. 2*B*). The monkey subjects could exhibit a preference for real faces (green line), an avoidance of real faces (or a preference for the unrealistic synthetic faces; blue line), an uncanny valley effect (a decreased preference for the realistic synthetic faces relative to the two other types; red line), an uncanny peak (a preference for the realistic synthetic face; purple line), and last, no preference at all (or general lack of interest, black line). Surprisingly, given these numerous possible outcomes, the monkeys consistently exhibited the uncanny valley effect (Fig. 3*A*), preferring to look at the unrealistic synthetic and real faces more than at the realistic synthetic faces. A Kruskal–Wallis nonparametric ANOVA for render-type revealed a significant overall effect based on the number of fixations [$\chi^2(2, n = 72) = 14.000$, $P = 0.001$]. A Mann–Whitney U test determined that the difference between unrealistic and realistic synthetic faces was significant ($z = -3.498$, $P < 0.001$), as was the difference between the realistic synthetic and real faces ($z = -2.806$, $P = 0.005$). There was no significant difference between unrealistic synthetic and real faces ($z = -0.877$, $P = 0.381$). Indeed, all five subjects displayed this pattern of looking preference (binomial test, $P = 0.03$; Fig. 3*B*). Although the numbers of fixation reveal that the subjects' visual behavior falls into the uncanny valley, it is possible that, despite their fewer fixations, they could be still looking longer at the realistic synthetic faces by increasing the fixation duration. We tested this possibility, and found that the mean fixation duration still revealed a significant effect [$\chi^2(2, n = 72) = 11.768$, $P = 0.003$] with subjects looking for the least amount of time toward the realistic synthetic face (versus unrealistic synthetic face, $z = -2.412$, $P = 0.016$; versus real face, $z = -3.402$, $P = 0.001$) (Fig. 3*C*). There was no significant fixation duration difference between unrealistic synthetic and real faces ($z = -0.474$, $P = 0.635$).

Although it has not been directly tested in any species (including humans), one of the predictions of the uncanny valley effect is that it should be stronger for dynamic versus static faces, that the 'valley' should be deeper for dynamic faces (Fig. 1) (1). Our data show that monkeys looked longer at the dynamic versus static faces overall ($z = -3.849$, $P < 0.001$; Fig. 4*A*), and that the uncanny valley effect was slightly more robust for dynamic faces. For dynamic faces, there was a significant effect for render type [$\chi^2(2, n = 30) = 7.543$, $P = 0.023$], with a significant difference between unrealistic versus realistic synthetic faces ($z = -2.570$, $P = 0.01$) and a marginally significant difference between realistic synthetic versus real faces ($z = -1.814$, $P = 0.07$) (Fig. 4*B*). For static faces, again there was a significant effect for

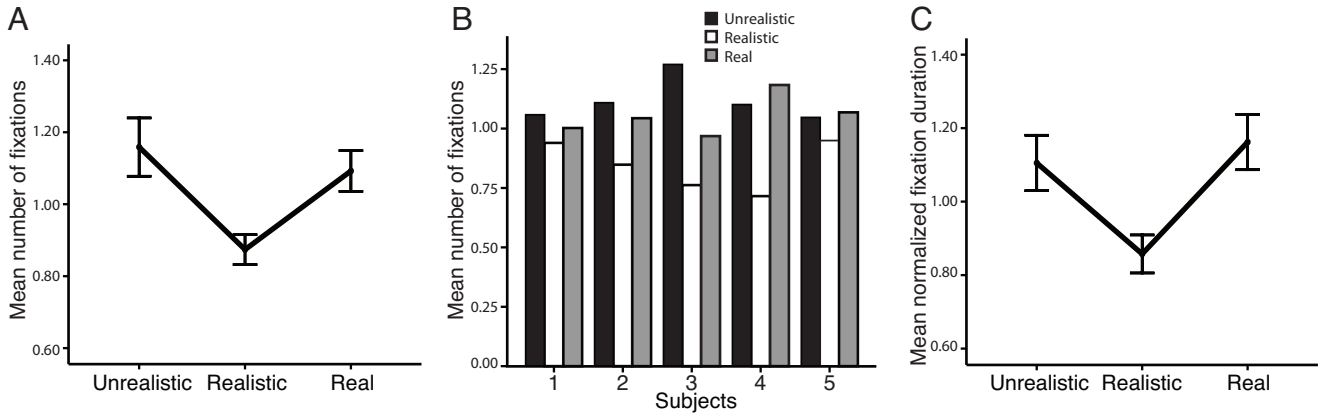

Fig. 3. Looking preferences of monkeys to different render-types. (*A*) Total number of mean-normalized fixations on each of the render-types. (*B*) All five monkey subjects show a looking pattern that reveals a greater number of fixations occur on the real agent and unrealistic synthetic agent than on the realistic synthetic agent. The *y* axis depicts mean-normalized number of fixations. (*C*) Total duration of mean-normalized fixations on each of the render-types. Error bars indicate SEM.

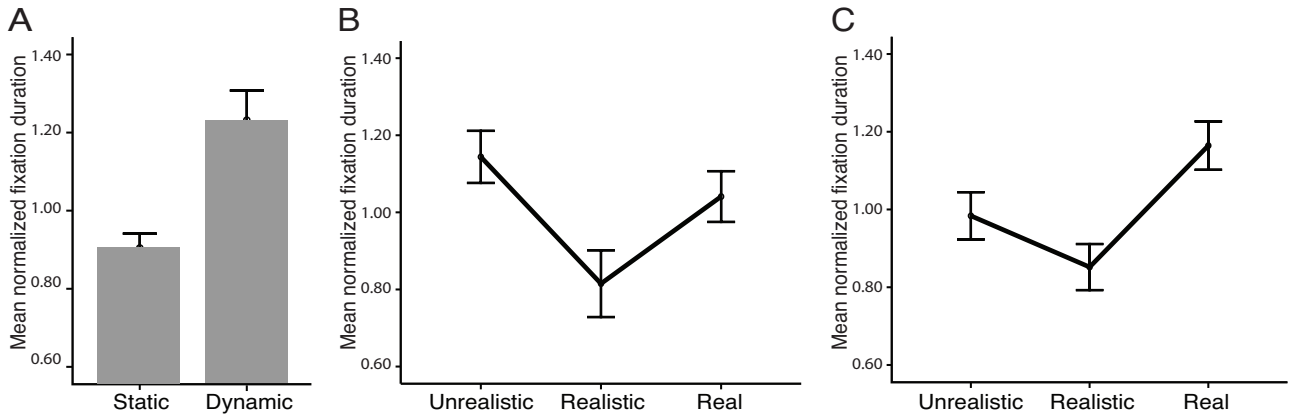

Fig. 4. Responses to static versus dynamic render-types. (*A*) Comparison of mean-normalized duration fixated for static and dynamic representations. (*B*) Total duration of mean-normalized fixations on each of the render-types within the dynamic stimuli data subset. (*C*) Total duration of mean-normalized fixations on each of the render-types within the static stimuli data subset. Error bars indicate SEM.

render type [$\chi^2(2, n = 42) = 9.444, P = 0.009$], but with no significant difference between unrealistic and realistic synthetic faces ($z = -1.287$, $P = 0.198$), and a significant difference between realistic synthetic and real faces ($z = -2.941, P = 0.003$) (Fig. 4*C*). In neither case (dynamic versus static) was the fixation duration significantly different between the unrealistic synthetic and real faces.

The uncanny valley effect in monkeys was not driven by expression type (coo versus scream versus neutral faces) as measured by either fixation duration [$\chi^2(2, n = 72) = 0.386, P = 0.825$] or by the number of fixations [$\chi^2(2, n = 72) = 0.221, P = 0.896$]. Differences in luminance among the stimuli also did not drive the effect. For the average screen luminance, stimuli were ranked according to their brightness, and there were no significant effects for fixation duration [$\chi^2(14, n = 72) = 21.852, P = 0.082$] or number of fixations [$\chi^2(14, n = 72) = 11.977, P = 0.608$].

In summary, out of five possible patterns of looking preferences toward faces with different levels realism (Fig. 2*B*), monkeys exhibited one pattern consistently: They preferred to look at unrealistic synthetic faces and real faces more than to realistic synthetic faces. The visual behavior of monkeys falls into the uncanny valley just the same as human visual behavior (2, 3). Thus, these data demonstrate that the uncanny valley effect is not unique to humans, and that evolutionary hypotheses regarding its origins are tenable. For example, although we cannot determine whether our monkey subjects find the realistic synthetic faces less attractive than real faces, we do know that many of the facial features that drive attractiveness in humans, such as facial coloration, may also influence the visual preferences of monkeys (13–16). What cannot be discerned in our experiment is whether the monkeys are experiencing disgust or fear (or aversion more generally) when they look at the realistic synthetic faces. It is possible that the monkeys find both the unrealistic synthetic faces and real faces more attractive than the realistic synthetic faces. However, given the copious amounts of anecdotal evidence from humans, and the more recent empirical studies supporting the uncanny valley effect in humans (2, 3), it seems parsimonious to conclude that monkeys are also experiencing at least some of the same emotions. To support this notion, future experiments could incorporate somatic markers, such as skin conductance responses, pupil dilation, or facial electromyography, to measure the emotional responses of monkeys to the differently rendered faces from monkeys (17–20). Neurophysiological and neuroimaging approaches could also be illuminating in this regard. For example, do synthetic faces paired with real voices elicit the same type and magnitude of multisensory integration and cortical interactions as real faces (21–23)?

As in humans (6), if monkeys exhibit an uncanny valley of face perception, then it may be because their brains are, to a greater degree than with unrealistic synthetic agents, processing the realistic synthetic agents as conspecifics–conspecifics that elicit, and fail to live up to, certain expectations regarding how they should look and act. Importantly, it is not the increased realism that elicits the uncanny valley effect, but rather that the increased realism lowers the tolerance for abnormalities (24). Understanding the features that exceed this tolerance threshold is a topic of intense investigation, particularly for those investigators interested in human-robot interactions and computer-animated avatar technology (4, 7). One likely possibility is that the computer-animated faces simply cannot capture all of the rapid and subtle movements of the face, and thus, there is an expectancy violation when looking at realistically-rendered faces (6). Although facial dynamics are certainly important to primate social perception and neurobiology (25), our data suggest that they do not seem to be the sole driving force or even a prominent one [contra Mori (1)]. Although monkeys looked longer at dynamic versus static faces, the differences between the uncanny valley effects between the two conditions were marginal at best. Data from human subjects suggest that various static features can reliably elicit or ameliorate the uncanny valley effect (2, 3, 7). These features include skin texture, the distance between facial features, and the size of various facial features. The same is likely to be true for monkeys.

That monkeys exhibit the uncanny valley effect suggests that the realistic synthetic agents are capable of eliciting monkey-directed expectations as though they were conspecifics. How such expectations are constructed is not known, but it is likely driven by social experience in both monkeys and humans (26–29). With further development, synthetic agents (both monkey- and human-like) can be used as a fully controllable actor in experiments investigating the neurobiology of dyadic social interactions (30–33). Such experiments will allow neural investigations of controlled, real-time dyadic interactions that obviate the need for the highly artificial trial-by-trial structure typical of neurophysiological studies of social processes.

Materials and Methods

Subjects. The study comprised five male long-tailed macaques (*M. fascicularis*). The subjects were born in captivity and socially pair-housed indoors. All experimental procedures were in compliance with the local authorities, National Institutes of Health guidelines, and the Princeton University Institutional Animal Care and Use Committee.

Stimuli. The visual stimuli consisted of digital video clips in Xvid (www.xvid.org) format featuring a vocalizing macaque monkey and two computer-generated (CG) primates differing in realism. The animated CG stimuli were generated using 3D Studio Max 8 (Autodesk) and were extensively modified from a stock model made available by DAZ Productions. The CG stimuli were modified to be unique with respect to their texture quality and polygon count. The realistic synthetic face featured a realistic texture and high polygon count to give it a rich, smooth, true-to-life appearance. The other CG stimulus (unrealistic synthetic face) featured no texture, was rendered in grayscale with salient red pupils, and used approximately one-third as many polygons as the high-quality version in an effort to give it an industrial, robot-like appearance. This reduction in the number of polygons represented the absolute minimum threshold whereby the model retained an appearance that was morphologically and kinesthetically similar to both the real monkey face and the realistic synthetic face.

Each monkey-type featured three facial expressions with corresponding audio tracks in two of them. The expressions used were a neutral face, a scream vocalization, and a coo vocalization. The neutral faces did not include audio tracks and were static images only. For both the scream and coo faces, a static version and a dynamic version were generated. The dynamic versions were digital videos, 2 s in duration with a frame rate of 30 frames per second. To control for audio, the same two audio tracks from the rhesus macaque scream and coo vocalizations were dubbed over the corresponding CG stimuli with Adobe Premiere Pro 1.5 at 32 kHz, 16-bit mono. CG stimuli were deliberately constructed to precisely match the auditory component from beginning to end. The static versions of these faces were frames extracted from the dynamic videos that represented the maximal expression and did not include an audio track.

Behavioral Apparatus and Paradigm. Experiments were conducted in a sound attenuating booth. The monkey sat in a primate chair fixed 65 cm opposite a 17-inch LCD color monitor with a 1280 × 1024 screen resolution and 60 Hz refresh rate. The 1280 × 1024 screen subtended a visual angle of 29°19′12″ (29.32°) horizontally and 23°27′36″ (23.46°) vertically. All stimuli were centrally located on the screen and occupied an area of 800 × 600 pixels, subtending a visual angle of 18°21′36″ (18.36°) horizontally and 14°27′36″ (14.46°) vertically. Eye movement data were captured with an infrared eye tracker, ASL Eye-Tracker 6000 (R6/Remote Optics), which featured a maximum resolution accurate to ≈3 mm at a distance of 65 cm (www.a-s-l.com). Because of the known location and accompanying fixations, the central fixation point used to begin trials was used as a standard candle to gauge the true calibrated resolution. Due to movement or initial miscalibrations, true resolution averaged ≈2°18′ (2.3°) horizontally and 2°7′12″ (2.12°) vertically. The monkeys did not have head-posts, and thus, to facilitate eye tracking, their heads were positioned between two Plexiglas pieces that limited their head rotation (34).

On entering the experimental chamber, the monkeys were first engaged in a two-point opposing-corner calibration routine. On successful calibration, monkeys were then required to fixate for 100 ms on the screen center to begin a session. Sessions comprised 3 trial-sets for each of the 15 stimuli; a total of 45 trials were conducted per session. All stimuli were randomized within a trial set. A trial contained a 2-s stimulus display followed by a 4-s intertrial interval that incorporated random colorful scenic photographs or farm animals. Three sessions were conducted in total resulting in nine trials per stimulus condition. Monkeys were continuously rewarded with apple juice at a flow rate of 20 mL/min for looking anywhere on screen at any given time during stimulus presentation to give them incentive to fixate on the screen and not elsewhere around the booth.

Because the subjects are being tested for their spontaneous reactions to these stimuli, we could not parametrically test ''realism'' along different dimensions such as polygon count, shading, or the distance between various facial features. In the ''looking time'' paradigm that we used here, monkeys quickly habituated to the repeated presentations of the different face types. Thus, it is not possible to vary the degree realism along different dimensions (as has been done in human studies; see refs. 3, 24), because it would require too many trials.

Data Analysis. Eye movement data were analyzed for fixations with Eyenal; a software tool provided by ASL for use with the Eye-Tracking system. Eyenal calculated fixations were evaluated as periods in which eye movement dispersion did not exceed 2° per 100 ms. The mean position of the measurements taken within a fixation period served as the overall eye position. MATLAB (Mathworks) was used to plot the extracted fixation locations onto the corresponding stimuli videos, as well as tally their number and duration. To compensate for peripheral vision, subject movement, and/or small miscalculations in calibration, fixations that fell within the total 1280 × 1024 screen were counted and included in the analyses performed with SPSS. All fixations occurring outside the screen area were discarded.

To analyze whether screen luminance had a significant effect on the number of fixations, relative screen luminance was systematically calculated. Using MATLAB, all frames from a given stimulus video were first converted from a triangular red, green, blue (RGB) color space into an equivalent hue, saturation, brightness (HSB) color space. The brightness values across all pixels for a given stimulus were then averaged to obtain a relative screen luminance.

For the 15 stimuli, the fixated duration and number of fixations on the screen across nine trials was summated. The resulting totals were normalized to the within-subject mean. The original data were also broken down into static and dynamic conditions, and separately analyzed for effects. As with the first analysis, fixation data were summated across nine trials and normalized to the within-subject means for static and dynamic presentations. Last, the data were broken down by facial expression into three conditions (neutral, coo, and scream), and separately analyzed for effects. As with the first analysis, fixation data were summated across nine trials and normalized to the within-subject means for neutral, coo, and scream presentations.

ACKNOWLEDGMENTS. We thank Karl MacDorman, Michael Platt, and Laurie Santos for their comments on a previous version of this manuscript. This work supported by National Science Foundation CAREER Award BCS-0547760.

1. Mori M (1970) The uncanny valley. *Energy* 7:33–35 (in Japanese).
2. MacDorman KF, Green RD, Ho C-C, Koch CT (2009) Too real for comfort? Uncanny responses to computer generated faces. *Comput Hum Behav* 25:695–710.
3. Seyama J, Nagayama RS (2007) The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence* 16:337–351.
4. Geller T (2008) Overcoming the uncanny valley. *IEEE Comput Graph Appl* 28:11–17.
5. Kahn PH, et al. (2007) What is human? Toward psychological benchmarks in the field of human-robot interaction. *Interact Stud* 8:363–390.
6. MacDorman KF, Ishiguro H (2006) The uncanny advantage of using androids in cognitive and social science research. *Interact Stud* 7:297–337.
7. Walters ML, Syrdal DS, Dautenhaun K, te Boekhorst R, Koay KL (2008) Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in a an attention-seeking home scenario for a robot companion. *Auton Robot* 24:159–178.
8. Rozin P, Fallon AE (1987) A perspective on disgust. *Psychol Rev* 94:23–41.
9. Jones BC, Little AC, Perrett DI (2004) When facial attractiveness is only skin deep. *Perception* 33:569–576.
10. Rhodes G, Tremewan T (1996) Averageness, exaggeration, and facial attractiveness. *Psychol Sci* 7:105–110.
11. Maner JK, Galliot MT, Rouby DA, Miller SL (2007) Can't take my eyes off you: Attentional adhesion to mates and rivals. *J Pers Soc Psychol* 93:389–401.
12. Humphrey NK (1974) Species and Individuals in Perceptual World of Monkeys. *Perception* 3:105–114.
13. Gerald MS, Waitt C, Little AC, Kraiselburd E (2007) Females pay attention to female secondary sexual color: An experimental study in Macaca mulatta. *Int J Primatol* 28:1–7.
14. Waitt C, Gerald MS, Little AC, Kraiselburd E (2006) Selective attention toward female secondary sexual color in male rhesus macaques. *Am J Primatol* 68:738–744.
15. Waitt C, et al. (2003) Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice. *Proc R Soc London Ser B Bio* 270:S144–S146.
16. Ghazanfar AA, Santos LR (2004) Primate brains in the wild: The sensory bases for social interactions. *Nat Rev Neurosci* 5:603–616.
17. Hoffman KL, Gothard KM, Schmid MC, Logothetis NK (2007) Facial-expression and gaze-selective responses in the monkey amygdala. *Curr Biol* 17:766–772.
18. Laine CM, Spitler KM, Mosher CP, Gothard KM (2009) Behavioral Triggers of Skin Conductance Responses and Their Neural Correlates in the Primate Amygdala. *J Neurophysiol* 101:1749–1754.
19. Zangehenpour S, Ghazanfar AA, Lewkowicz DJ, Zatorre RJ (2008) Heterochrony and cross-species intersensory matching by infant vervet monkeys. *PLoS ONE* 4:e4302.
20. Waller BM, Parr LA, Gothard KM, Burrows AM, Fuglevand AJ (2008) Mapping the contribution of single muscles to facial movements in the rhesus macaque. *Physiol Behav* 95:93–100.
21. Chandrasekaran C, Ghazanfar AA (2009) Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. *J Neurophysiol* 101:773–788.
22. Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the Superior Temporal Sulcus and Auditory Cortex Mediate Dynamic Face/Voice Integration in Rhesus Monkeys. *J Neurosci* 28:4457–4469.
23. Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004–5012.
24. Green RD, MacDorman KF, Hoa C-C, Vasudevana S (2008) Sensitivity to the proportions of faces that vary in human likeness. *Comput Hum Behav* 24:2456–2474.
25. Shepherd SV, Ghazanfar AA (2009) Engaging neocortical networks with dynamic faces. *Dynamic Faces: Insights from Experiments and Computation*, eds Giese M, Curio C, Buelthoff H (MIT Press, Cambridge, MA), in press.
26. Lewkowicz DJ, Ghazanfar AA (2006) The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci USA* 103:6771–6774.

27. Sugita Y (2008) Face perception in monkeys reared with no exposure to faces. (*Proc Natl Acad Sci USA* 105:394–398.
28. Langlois JH, et al. (1987) Infant preferences for attractive faces: Rudiments of a stereotype. *Dev Psychol* 23:363–369.
29. Lewkowicz DJ, Ghazanfar AA (2009) The emergence of multisensory systems through perceptual narrowing. *Trends Cognit Sci*, doi: 10.1016/j.tics.2009.08.004.
30. Gong L (2008) How social is social response to computers? The function of the degree of anthropomorphism in computer representations. *Comput Hum Behav* 24:1494–1509.
31. Krach S, et al. (2008) Can machines think? Interaction and perspective-taking with robots investigated via fMRI. *PLoS ONE* 3:e2597.
32. Wheatley T, Milleville SC, Martin A (2007) Understanding animate agents: Distinct roles for the social network and mirror system. *Psychol Sci* 18:469–474.
33. Moser E, et al. (2007) Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *J Neurosci Methods* 161:126–133.
34. Nemanic S, Alvarado MC, Bachevalier J (2004) The hippocampal/parahippocampal regions and recognition memory: Insights from visual paired comparison versus object-delayed nonmatching in monkeys. *J Neurosci* 24:2013–2026.