AUTHOR'S REPLY TO COMMENTARIES

Opening Pandora's Box

Reply to commentaries on "The uncanny advantage of using androids in social and cognitive science research"

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

Androids have the potential to reinvigorate the social and cognitive sciences — both by serving as an experimental apparatus for evaluating hypotheses about human interaction and as a testing ground for cognitive models. Unlike other robotics techniques, androids can illuminate how interaction draws on human appearance and behavior. When cognitive models are implemented in androids, feelings associated with the uncanny valley provide heightened feedback for diagnosing flaws in the models during human–android interaction. This enables a detailed examination of real-time factors in human social interaction.

Not only can android science inform us about human beings, but it can also contribute to a methodology for creating interactive robots and a set of principles for their design. By doing this, android science can help us devise a new kind of interface. Since our expressive bodies and perceptual and motor systems have co-evolved to work together, it seems natural for robot engineers to exploit this by building androids, rather than hoping for people to gradually adapt themselves to mechanical-looking robots. In the longer term, androids may prove to be a useful tool for understanding social learning, interpersonal relationships, and how human brains and bodies turn themselves into persons (MacDorman & Cowley, 2006). Of course, there are many ways to investigate human perception and interaction and to explore the potential for interactive robotics. Android science is only one of them. Although the uncanny valley plays a special role in android science, the nature of the phenomenon should rightly be investigated by other approaches too.

Interaction Studies 7:3 (2006), **361–368**.
ISSN 1572–0373 / E-ISSN 1572–0381 © John Benjamins Publishing Company

The Uncanny Valley

It is important to separate the study of the uncanny valley as a perceptual phenomenon from its special role in android science and its wider impact on society and the economy with regard to computer animation, video games, and the marketing of entertainment robots and other products. Let us first consider its role in android science. Granted that a cognitive model is implemented in a human-looking robot, the uncanny valley provides heightened feedback concerning how the model deviates from human norms of behavior. This diagnostic function enables the progressive refinement of the model. Thus, in android science the uncanny valley contributes *positively* to our understanding of human beings, while allowing us to build robots that act more human.

However, Ziemke and Lindblom (2006) illustrate the important methodological issues android science will face in adopting a top-down, synthetic approach to investigating the uncanny valley. For example, they cite the difficulty of distinguishing differences in human response, and especially subconscious response, caused by a lack of humanness from those caused by other factors. They also note a credit assignment problem when working with a system as complex as an android in determining which of its subsystems was responsible for causing a difference in human response. However, the problem of tracing causes to internal mechanisms is a general problem for cognitive neuroscience, and the problem of tracing causes to perceived behavior is a general problem for the social sciences.[1] Therefore, these issues are unlikely to impede progress in android science any more than the progress of these disciplines. Androids at least have the advantage of possessing internal workings that are easily accessed. In addition, we do recognize the value of bottom-up approaches to studying the uncanny valley, and we have pursued them in our own research. The main focus of the top-down approach is on face-to-face interaction.

Ziemke and Lindblom (2006) represent the new consensus that bodies matter: interaction cannot be characterized by disembodied, decontextualized models that define behavior to be the output of inner information processing. We are sympathetic to Ziemke and Lindblom's criticism of strong AI and their claim that the Turing test is an inadequate test of intelligence, which is why we proposed a communion game in the spirit of Turing's argument for learning robots but without any strong AI claims (cf. Cowley & MacDorman, 1995). We also agree that it would be fruitful to investigate intelligence through the emergent processes of embodied development (Cowley & MacDorman, 2006; MacDorman & Cowley, 2006).

Chaminade and Hodgins (2006) argue for the benefits of using a broad range of stimuli, from points-of-light displays to mechanical-looking robots, to tackle human perception in general and the uncanny valley in particular, and we concur with them. We do *not* "posit the use of androids as the only key that will open the door to the secrets of the [uncanny] valley" (Cañamero, 2006). Near-human forms might sometimes seem eerie because they are eliciting a cognitive mechanism that evolved for another purpose:[2] for example, to perceive as attractive potential mates who are fertile (Etcoff, 1999) or to perceive as disgusting entities that are likely to transmit disease (Rozin, 1987). Such cognitive mechanisms, which evolved long before robots existed, can be studied by means of the *same* stimuli they evolved to detect.[3] But the feeling of eeriness associated with the uncanny valley could instead be caused by entities that 'fall through the cracks' of these and other cognitive mechanisms — by entities that lie on category boundaries like the boundary between *human* and *machine* (Ramey, 2005). This boundary effect is likely to be a general phenomenon that can be approached in many different ways. However, since most people lack experience with androids, casual attributions about their eeriness could be nothing more than a common report. Given the diversity of explanations of the uncanny valley and the range of other stimuli found to be eerie, the valley may well be a hodgepodge of factors (MacDorman & Ishiguro, 2006) — in other words, several "doors" requiring perhaps even more "keys" to open them.

Kosloff and Greenberg (2006) express concern that the uncanny valley could make androids less acceptable to society — for example, by reminding people of human corpses. We make no judgment about the *moral* value of building robot products intended to be or not to be uncanny. Their *economic* value will be assessed by product marketers, and it is worth noting that a niche has already been found for the 'scary cute' genre of dolls in Japan. It is also clear from the popularity of the horror and thriller genres that many people actively seek out the uncanny. David Hanson has intentionally challenged tastes by exhibiting the humanlike robot heads he has built with motors and wires exposed. People who have interacted with these heads have expressed disappointment about them ceasing to seem strange once they get used to them. Therefore, it seems the uncanny is not necessarily a bad thing.

Bottom-up and Top-down Approaches in Android Science

Chaminade and Hodgins (2006) note that both the top-down approach we have advocated and bottom-up approaches are valuable to research in cognitive psychology and cognitive neuroscience. The approaches are complementary, and neither can resolve everything. The approach of using functional magnetic resonance imaging (fMRI) to investigate the brain's mirror system is highly relevant to understanding social interaction and other issues we want to resolve.

However, we should separate issues related to android science into several tracts. One concerns how neural systems including the mirror system respond to particular movements. These systems contribute to our predisposition to anthropomorphize by relating the movements of others to our own bodily movements. But research using points-of-light displays or mechanical-looking robots does not explain why a humanlike body and face affect us so strongly (Cole, 2001; Hirai & Hiraki, 2006). Previous work in human–robot interaction has ignored, or failed to control for, the influence of human appearance and behavior. Understanding their influence is important to understanding the brain functions under study. For example, preliminary results from ongoing fMRI experiments prepared at Hiroshi Ishiguro's lab at Osaka University and performed at the Department of Cognitive Science, UCSD, suggest that, when performing an identical reaching movement, an android is able to enlist the mirror system to a greater extent with its skin on than with its mechanical underpinnings exposed. This finding demonstrates the importance of appearance. It also illustrates how fruitful research can combine brain imaging methods with androids and other kinds of stimuli. In this sense, the approach of using androids is not strictly top-down but can resemble some of the bottom-up approaches mentioned by the commentators.

Another major tract of research in android science concerns the total evaluation of a robot as an interactive agent. How can we design robots that excel at interacting with people? Bottom-up approaches have generated many findings, but we do not know how to integrate them in a machine that is capable of becoming anything remotely like a person (MacDorman & Cowley, 2006). Integration by design is difficult and, as Ziemke and Lindblom (2006) indicate, may not be appropriate. But even for control systems that self-organize, androids provide an essential testing ground to investigate how various cognitive functions develop in an agent in a well-orchestrated way and how they scale-up to human interaction. Starting with the most humanlike agents we can build offers the greatest control over experimental conditions by eliminating many of the effects of non-human appearance and behavior.

One of our goals in building androids is to develop a general methodology for creating interactive robots and to find effective design principles, so it is important to involve other kinds of robots and even non-robot entities. However, the problem lies with how to evaluate mechanical-looking robots. What can be learned from them? Because these robots do not look human, even if they reproduced human reactions perfectly, we could not evaluate the *human likeness* of their behavior during interaction with people because they would not elicit *human-directed* response.[4] This makes it doubtful that studies on interaction with mechanical-looking robots could generalize to human beings. But as we build more human-looking androids, we can ignore many factors, like the influence of appearance, and focus on the target problem.

If you want to understand how people interact with a mechanical-looking robot, you use a mechanical-looking robot. If you want to understand how people interact with each other, a mechanical-looking robot is not enough. Since we want to understand human beings, we build androids. We intend to develop a science of human–*human* interaction, not a science of human–Kismet (or human–Feelix, etc.) interaction, although these other sciences would be of interest to the people who built these robots. But from the perspective of those who set government funding objectives, a science of human interaction would be better positioned to compete with such other worthy priorities as curing cancer and AIDS.

Given our focus on closely coordinated full-bodied interaction, we cannot accept that all media are essentially humanlike. The importance of Reeves and Nass's (1996) work is that they showed how people can respond socially to their prior interactions with computers and other media. But computers already have a relatively stable relationship to people. Computer software is task-oriented, and there are few ambiguities concerning its intended use. However, we do not know what the best design of an interactive robot is or even what its role should be. This is another reason to focus on developing androids. Androids avoid the design ambiguities of other kinds of interactive robots because the design policy is to build an artificial human being — a person (MacDorman & Cowley, 2006).

The goal of building androids allows us to go beyond merely *equating* some aspects of computer or robot-directed response with human-directed response. We can begin to explore real-time human interaction by implementing responses in robots that leverage on the expressive power of the human form — in robots whose form subconsciously and immediately tells us how to interact with them. Contrary to this viewpoint, Kosloff and Greenberg (2006) state that the same arguments we used against human–humanoid interaction concerning its inability to generalize to people can be applied to human–android interaction. They note that the research we cite "shows that interactions with robots and androids elicit quite different reactions than those elicited in interaction with humans" (Kosloff & Greenberg, 2006). But that just means we are not *there* yet, which is to say that androids are not human enough yet. Our androids can fool most people into thinking they are human for two seconds, not five or ten minutes, which is the length of some of our cited experiments. As Chaminade and Hodgins (2006) point out, an enormous amount of work must be done to create androids that can pass as human in extended interactions, so clearly we are only at the beginning stages.

But even with our current prototypes, Japanese participants made eye contact by looking at the right eye of the android just as they did when interacting with a person. This is totally different from their fixation pattern with a mechanical-looking humanoid robot, in which case their eyes roamed all over the robot's face and torso. In the experiment on gaze under cognitive load, Japanese participants already show the same modesty with their eyes by looking down when interacting with the android as when interacting with a person *if* they were told the android was under human control.[5] Since one of us, Ishiguro, now has an android twin to control by telepresence, the illusion that there is a human mind behind the machine may not be so hard to maintain.

If robots are to interact with people and assume human roles, is it important for them to look human? The commentators rightly note that the answer will depend on the context and cannot be determined without experiments. But in the meantime, we can make some educated guesses based on examples from human interaction. So let us turn the question around and ask whether it is important for human beings to look human. Evidence suggests that it is. In "Empathy Needs a Face," Jonathan Cole (2001) describes interviews with patients who have reduced facial expressiveness caused by Moebius Syndrome or Parkinson's disease. He discovered that after the disease's onset even gregarious extroverts can seem like sullen introverts. Patients with these conditions find it hard to capture the interest of others or to join in a conversation. If facial movements that violate human norms make it hard for people to interact with other people, how much harder would they make it for simple robots that do not even look human? Although people purport to feel empathy for simple robots (Cañamero, 2006),[6] the plight of patients with Moebius Syndrome or Parkinson's disease demonstrates how unhelpful that sort of data is to understanding real-time social interaction.

We run against the limitations of simple robots, when we try to study how gaze, gesture, speech, and facial expressions come together during human interaction in ways that are closely coordinated, multimodal, and highly contingent. The cameras of a simple robot look nothing like eyes, for a start, so people make little eye contact with them. So we may add a pair of humanlike eyes to the robot. But we find this is not enough. We would like to study how gaze is coordinated with human touch and gesture, but the robot has nothing like human limbs or skin, so we add limbs and skin. But again, we find this is not enough. And so we decide to build an android. And when we find that our android prototype is still not enough — the results are not identical to those of human–human interaction — what do we do? Do we improve the android, or do we go back to using the simple robot? It would seem the path to understanding human interaction invariably leads to *improved* androids.

Notes

[1] Consider, for example, Quine's (1960, 1977) argument concerning the indeterminacy of translation. In addition, credit assignment is further complicated by the fact that human behavior is to some extent nondeterministic.

[2] Although human beings have always had to discriminate among their conspecifics and other species, only in science fiction are android detectors essential to the survival of the species.

[3] Homo sapiens may have perceived an uncanny valley in the ascent of their hominid ancestors. If their Neanderthal 'cousins' were unlucky enough to fall into the valley, this could have hastened their extinction.

[4] Simulated human beings suffer from the same limitation. If anything, the story of Greta Garbo's arrival at MGM (Cañamero, 2006) illustrates the need to build physical androids, because it shows how people's reaction to physical presence and screen presence can be so different.

[5] This experiment also shows that interaction is irreducible to behavior, because we do not simply respond to what a machine does since beliefs and norms also play a role in real-time responding (MacDorman et al., 2005).

[6] Braitenberg (1984) reports similar results with simple vehicles. Decades earlier Michotte (1962) and Heider and Simmel (1944) also found that human participants would attribute mental states to moving circles or other geometric shapes. Human participants may feel empathy for a little square on a screen being bullied by a big square. But there is only so much one can do with it, which is probably why cartoons have gone the direction of heightened human expressivity. Even Mickey Mouse — who is supposed to be a *mouse* after all — can walk, talk, fish, and type like a human. The drawing of the character has also come to look more like a human and less like a mouse over time.

References

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology.* Cambridge, Mass.: The MIT Press.

Cañamero, L. (2006). Did Garbo care about the uncanny valley? *Interaction Studies, 7*(3).

Chaminade, T. & Hodgins, J.K. (2006). Artificial agents in the social cognitive sciences. *Interaction Studies, 7*(3).

Cole, J. (2001). Empathy needs a face. *Journal of Consciousness Studies, 8*(5–7), 51–68.

Cowley, S.J. & MacDorman, K.F. (1995). Simulating conversations: The communion game. *AI & Society, 9*(3), 116–137.

Cowley, S.J. & MacDorman, K.F. (2006). What baboons, babies, and Tetris players tell us about interaction: A biosocial view of norm-based social learning. *Connection Science, 18*(3).

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

Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57,* 243–259.

Hirai, M. & Hiraki, K. (2006). Disappearance of inversion effect for walking animation with robotic appearance. In *The Proceedings of the Fifth International Conference of the Cognitive Science Society,* Toward Social Mechanisms of Android Science symposium. July 26, 2006. Vancouver, Canada.

Kosloff, S. & Greenberg, J. (2006).Android science by all means, but let's be canny about it! *Interaction Studies, 7*(3).

MacDorman, K., Minato, T., Shimada, M., Itakura, S., Cowley, S., and 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.

MacDorman, K. & Cowley, S.J. (2006). Single white robot seeks human companions for long-term relationships: A new benchmark for robot personhood. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication.* September 6–8, 2006. University of Hertfordshire, Hatfield, UK.

MacDorman, K.F. & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3).

Michotte, A. (1962). *The perception of causality.* Andover, Mass.: Methuen.

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

Quine, W.V.O. (1960). *Word and object.* Cambridge, Mass.: The MIT Press.

Quine, W.V.O. (1977). *Ontological relativity.* New York: Columbia University Press.

Ramey, C.H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In *Proceedings of the "Views of the Uncanny Valley" workshop, IEEE-RAS International Conference on Humanoid Robots,* Tsukuba, Japan.

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

Ziemke, T. & Lindblom, J. (2006). Some methodological issues in android science. *Interaction Studies, 7*(3).