An Interactive Virtual Reality Platform for Studying Embodied Social Interaction

Hui Zhang (huizhang@cs.indiana.edu)
Department of Computer Science; Indiana University
Chen Yu (chenyu@indiana.edu)
Department of Psychology and Brain Sciences; Indiana University
Linda B. Smith (smith4@indiana.edu)
Department of Psychology and Brain Sciences; Indiana University

Abstract

We present an interactive virtual reality platform for studying the role of embodied social interaction in the context of language learning. The virtual environment consists of virtual objects, a virtual table, and most importantly, a set of virtual students with different social-cognitive skills. Real users are asked to serve as language teachers and teach virtual learners object names. They can interact with virtual learners via gazing, pointing at and moving virtual objects as well as speech acts. Since both the virtual environment (what users see) and the virtual humans (whom users interact with) are controlled (pre-programmed), this provides a unique opportunity to study how real teachers perceive different social signals generated by virtual learners and how they adjust their behaviors accordingly. One primary result is that real people feel comfortable to interact with virtual humans in the virtual environment and treat them as social partners. Moreover, the platform allows us to record real people's multimodal behavioral data and analyze the data across individual participants to extract shared behavioral patterns. Overall, this work demonstrates the usefulness of virtual reality technologies in studying both human-human and human-machine social interactions.

Introduction

A better understanding of human-human interaction in language learning has long been a subject of fascination. Language learning is a social event between teachers and learners. Nonverbal communication, including body language, gaze, gesture, facial expression, is crucial for both smooth communication and effective learning. More specifically, body language signaled by a language teacher provides useful cues for a language learner to infer what the speaker intends to refer to in unknown (yet) language. For example, a deictic pointing action would single out one object from multiple ones in a natural scene and indicate the speaker's referential intentions [14]. Meanwhile, body language signaled by a language learner indicates his/her attentional state so that the language teacher can adjust behaviors accordingly to enhance interaction and learning. For instance, if the language teacher realizes that the learner is not engaged in the interaction, she would generate some actions to attract the learner's attention. On the other hand, if the learner is fully engaged, then the teacher would focus more on using body language to facilitate language learning (but not on engaging the language learner).

Although previous research demonstrates the importance of social cues in the laboratory environment [1], quantitative analyses of the role of social cues in real world is very difficult without interfering with the interaction itself. What is really needed is an approach to controlling dynamic interactions between the language teacher and the language learner. By doing so, we can decouple the social interactions between two agents and manipulate the parameters in the interaction dynamically and systematically in a well-controlled way. The present paper addresses this challenge by using state-of-art technologies in computer graphics and virtual reality.

In the past decade, applications of virtual reality (VR) technology have been rapidly developed with the advance of computer graphics software and hardware. Virtual Reality techniques provide a unique way to enable people to interact efficiently with 3D computerized characters in a computer-rendered environment in real time using their natural senses and skills. Recently there is a growing trend that VR can play an important role in basic research in a variety of disciplines including cognition[2], education [9, 4] and perception[11]. Among others, Jasso and Triesch presented a virtual reality platform for developing and evaluating embodied models of cognitive development in [6]. Turk et al.[13] introduced a paradigm for studying multimodal and nonverbal communication in collaborative virtual environment where a user's communication behaviors can be filtered and re-rendered in a VR environment to change the nature of social interaction.

In light of this, we present a new experimental paradigm that exploits VR technologies to decouple complex social interactions between two agents and to study the role of embodied social cues in language learning. Specifically, we hypothesize that naturalistic social influence can occur within immersive virtual environments as a function of two additive factors, behavioral realism and social presence. This paper takes the first steps towards this goal by designing and implementing a novel interactive virtual reality platform by asking real users to interact with virtual humans through various embodied social interactions. We report a case study of using this virtual reality platform with the evaluations of this platform in the context of a language learning task.

System Framework

Overview

We build virtual humans equipped (pre-programmed) with different kinds of social cognitive skills and ask real people to interact with virtual humans in a virtual environment.

Our VR interaction system consists of four components as shown in Figure 1:

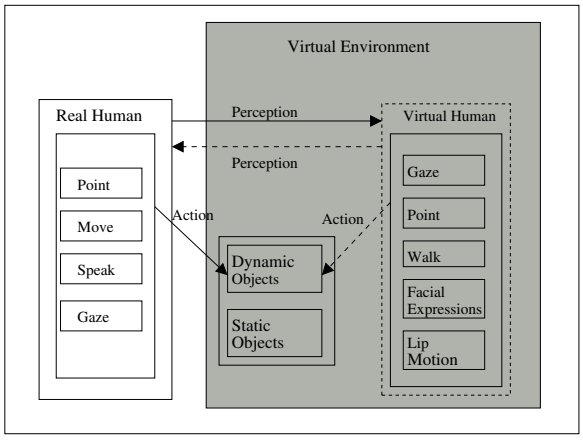

Figure 1: Overview of system architecture.

- A virtual environment includes a virtual laboratory with furniture and a set of virtual objects that real people can manipulate in real time via a touch screen mounted on a computer screen.
- Virtual humans can demonstrate different kinds of social skills and perform actions in the virtual environment.
- Multimodal interaction between virtual humans and real people includes speaking, eye contact, pointing at, gazing at and moving virtual objects.
- Data recording monitors and records a participant's body movements including pointing and moving actions on virtual objects, eye gaze, and speech acts in real time.

Building Virtual Humans

Appearance and Behavior One of the most important issues in our design is the "behavioral realism" of the virtual agents, which means that virtual humans should act and respond like a human, or in other words, they should be believable (see [12]), in both the physical actions of the agents themselves, and their social interactions with the human users.

The implementation at perceptual and motor levels is based on a human animation software package called *DI – Guy*, which is commercially available from Boston Dynamics Inc. It provides textured human characters with basic motor skills, such as standing, strolling, walking, running, sitting, etc. The actions of *DI – Guy* characters can be scripted manually using an interactive tool called *DI – GuyScenario*. The other option, which is the one we use, is based on *DI – Guy* SDK, allowing external C/C++ programs to control a character's basic motor repertoire. This SDK enables us to interface *DI – Guy* to our extensive, high-level attentional and cognitive control software. A sample virtual human is shown in Figure 2, suggesting that using DI-Guy can result in smooth and lifelike movements being generated automatically.

Attentional State In our system, virtual humans can be programmed to behave to be engaged or disengaged in the

Figure 2: Interacting with virtual agent. The virtual lady is paying attention to the attentional objects on the virtual table.

interaction. If she is engaged, she will generate a set of actions, such as following the visual attention of a real person, paying attention to the objects that the real person is manipulated, and showing positive facial expressions. If she is not engaged, she would look somewhere irrelevant the real person's actions and generate negative facial expressions. We suggest that eye gaze plays a pivotal role in face-to-face interaction. Therefore, the simulation of cognitive skill is based primarily on avatar's eye gaze and pointing models evident in the psychological literature, and our simulation takes advantage of many techniques that have been widely used in other avatar interfaces (see [8], [10], [3], [7] and [5]).

The highest level of our eye gaze model is based on transitions between the two states (i.e., gazing at attentional objects and gazing away from attentional objects). The transition is triggered primarily by the passing of time in the current state, which is controlled by the level of engagement. And when the virtual human is engaged in a social conversation, he should gaze at the attentional object the human user is attending to. A further example in Figure 3 shows various engagement levels on multiple agents can be modeled to simulate a teaching and learning environment.

Figure 3: Modeling students with different levels of engagement.

Interaction and Data Recording

As shown in Figure 1, a user and a virtual human can interact through multiple channels including pointing at and moving virtual objects via hands, gazing at objects via eyes, and generating facial expressions. We have developed a multimodal data recording program that collects participants' speech, gaze movement on the computer screen, and actions on the touch screen mounted the display computer monitor. Speech

signals were sampled at 8000Hz and the sampling rate of both actions on the touch screen and eye gaze is 60Hz.

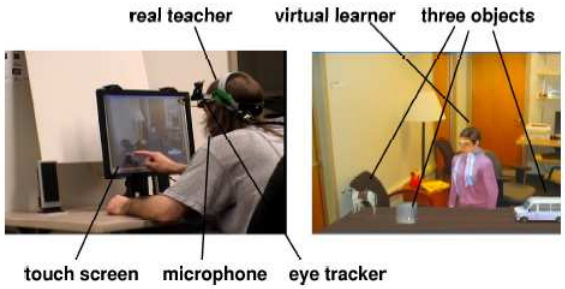

Figure 4: Left: a participant wearing an eye tracker and a microphone interacts with the virtual human in a virtual environment through a touch screen. Right: the VR scene consists of a virtual human and three objects on a table in each trial.

Platform Evaluation: Real humans teach virtual learners

As a first step to evaluate the usability of this platform, we designed an experiment in which real people were asked to teach virtual learners object names. We control the behaviors of virtual learners to create different learning situations, and measure how real people perceive the social-cognitive skills of different virtual people and how they adjust both their interactive behaviors and teaching strategies based on their perception of virtual learners.

Design and Procedure

As shown in Figure 4, real people were asked to teach virtual foreigners the names of several everyday objects. They were allowed to point to, gaze at and move those objects through a touch screen. There was no constraint about what they have to say or what they have to do. There were three conditions in this experiment wherein three virtual agents demonstrated different levels of engagement in interaction - engaged in 10%, 50% or 90% of total interaction time. When a virtual human is fully engaged in interaction, she would share visual attention with a real teacher by gazing at the object attended by a real teacher and generating positive facial expressions (e.g. smile, trust, etc.). While she is not engaged, she would look at somewhere else with negative facial expressions (e.g. sad, conniving, etc.). The objects attended by a real person are detected based on where he is looking as well as his actions on those objects through the touch screen. The attentional information is then sent to the virtual human so that she can switch her attention to the right objects in real time when she is in the engaged state.

We recruited 26 subjects who received course credits for participation. They were asked to interact with three virtual humans in total and one per condition. We randomly assigned the virtual humans to three levels of engagement, counterbalancing across participants.

There were six trials in each engagement condition and three virtual objects were introduced in each trial. Thus, participants needed to teach $3\times$ 6=18 objects in each condition and 54 objects in all of the three conditions. Whenever they thought that the virtual learner already acquired three object names in the current trial, they could move to the next trial. We recorded real people's behaviors in interaction including their pointing and moving actions, speech acts and eye gaze. Moreover, they were asked to complete questionnaires at the end of the experiment. The questionnaires measured social intelligence of three virtual learners. They were also asked to provide their estimates of the percentage of time the virtual humans followed the human teacher's attention.

Measure and Results

A 5-point Likert scale was used for a set of 10 questions in our questionnaire. Those questions focus on different aspects of participants' perception of the social-cognitive skills of three virtual humans:

- **Joint attention and eye contact** We measured how much the participants felt that eye movements of virtual humans were natural, social and friendly. A representative question contributed to this measure is "I felt that the agent did not look enough at me".
- **Social intelligence/engagement** We calculated a score to measure how much the participants felt that virtual learners were engaged during interaction (0-not engaged at all, 5-fully engaged). A representative question in this measure is "the agent and I interacted very smoothly".
- **Overall intelligence** We calculated a score to measure participants' estimates of virtual learners' intelligence. An example question used here is "the agent is smart".
- **Gaze time estimation**: Participants were also asked to estimate the amount of time (on a scale of 0 to 100 percent) that virtual humans paid attention to their behaviors.

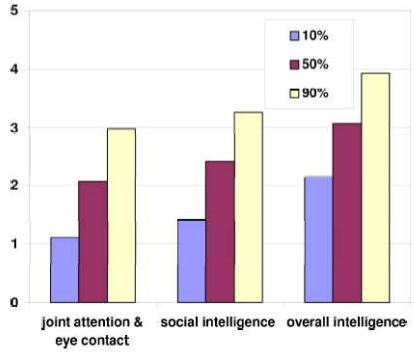

Figure 5: A comparison of participants' evaluation of three virtual humans.

Table 1: The estimated engagement times of virtual humans

	10%	50%	90%
gaze time	M= 22.50% SD= 22.10%	M=54.37% SD= 23.89%	M= 86% SD=16.1%

Figure 5 shows a comparison of the results of three virtual humans with different engagement levels. Clearly, participants

were aware of social behaviors of virtual humans and provided quite consistent estimates of their social sensitivities. Thus, the significant differences between three conditions are not surprising. We note that even when the virtual human almost fully engaged in interaction by following the real person's actions in 90% of the total time, most people were still not satisfied with the virtual human's social behaviors. Another observation is that they gave more credits to the high-level questions such as the overall intelligence of the virtual humans, but were less satisfied with more concrete issues, such as eye contact. This is true in all of the three conditions.

Table I shows the estimated times that virtual humans pay attention to participants' behaviors. Although the means of two out of three estimated times are close to 50% and 90% separately. Surprisingly, participants provided quite different estimates in all of three conditions. For instance, the low limit for the estimates in the 10% condition is 0%, indicating that some participants significantly overestimated the virtual human's engagement time. Meanwhile, some of them underestimate the times in the 90% condition as well. Further investigation is needed to explain this observation.

The purpose of these measures is to investigate whether the participants believe that they have been interacting with the representation of a real other (i.e., "social presence"). According to our experiments, social influence that occurs in the interacted virtual reality is accepted by the real participants. Our investigation shows that as far as those primitive actions generated by virtual humans look realistic, real people would treat them as social partners and are willing to interact with them.

Conclusion

Compared with using a real robot in a real environment, virtual humans are easy to implement and use mainly because we can neglect low-level technical problems, such as motor control of joint angles, which perfectly matches our research purposes. We are most interested in high-level social-cognitive skills in language learning. We attempt to answer how the behavioral-level actions, such as gazing and pointing, generated from both a language teacher and a language learner, are dynamically coupled in real time to create the social learning environment, and how the language learner appreciates those social cues signaled by the teacher. Moreover, the virtual platform has several special advantages in the study of social interaction: (1) Various virtual environments can be easily created and we can dynamically change or switch between different virtual scenes easily during an experiment; (2) the degree to fully control both virtual humans' behaviors and the virtual environment that real users and virtual humans share cannot be achieved with neither real robots nor human experimenters, which allows us to systematically study what aspects of the social environment are crucial for learning; and (3) we can easily maintain the consistency of the experimental environment and perfectly reproduce the experiments across multiple participants.

In summary, the present study proposes and implements a new experimental paradigm to study learning from multimodal interaction. We build virtual humans and control their behaviors to create different social partners that real people interacted with. We measured how well real people interact with virtual humans and how they shape their behaviors to adapt to different social-cognitive skills that virtual humans possess. We found that real people treat virtual humans as social partners when they interact with them, suggesting that we can further apply this experimental setup to create different interaction conditions by systematically manipulating the virtual human's behaviors.

References

[1] D.A. Baldwin. Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, 29:832–843, 1993.

[2] D.H. Ballard, M. M. Hayhoe, P.K. Pook, and R. P. Rao. Deictic codes for the embodiment of cognition. *Behavioural and Brain Science*, 1996.

[3] A. Colburn, M. Cohen, and S. Drucker. The role of eye gaze in avatar mediated conversational interfaces, 2000.

[4] S. D. Craig, B. Gholson, and D. M. Driscoll. Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94(2):428–434, June 2002.

[5] M. Garau, M. Slater, S. Bee, and M. A. Sasse. The impact of eye gaze on communication using humanoid avatars. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 309–316, New York, NY, USA, 2001. ACM Press.

[6] H. Jasso and J. Triesch. A virtual reality platform for modeling cognitive development. In G. Palm and S. Wermter, editors, *Biomimetic neural learning for intelligent robots*. Springer, 2005.

[7] K.F. MacDorman, T. Minato, M. Shimada, S. Itakura, S. Cowley, and H. Ishiguro. Assessing human likeness by eye contact in an android testbed. In *Proceedings of the XXVII Annual Meeting of the Cognitive Science*, page 226, Stresa, Italy, 2005.

[8] C. Peters and C. O'Sullivan. Bottom-up visual attention for virtual human animation. In *CASA*, pages 111–117, 2003.

[9] Jeff Rickel and W. Lewis Johnson. Task-oriented collaboration with embodied agents in virtual worlds. pages 95–122, 2000.

[10] T. Rist, M. Schmitt, C. Pelachaud, and M. Bilvi. Towards a simulation of conversations with expressive embodied speakers and listeners. In *CASA*, pages 5–10, 2003.

[11] C. S. Sahm, S. H. Creem-Regehr, W. B. Thompson, and P. Willemsen. Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Trans. Appl. Percept.*, 2(1):35–45, 2005.

[12] N. M. Thalmann, H. Kim, A. Egges, and S. Garchery. Believability and interaction in virtual worlds. In *MMM*, pages 2–9, 2005.

[13] M. Turk, J. Bailenson, A. Beall, J. Blascovich, and R. Guadagno. Multimodal transformed social interaction. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 46–52, New York, NY, USA, 2004. ACM Press.

[14] C. Yu, D. H. Ballard, and R. N. Aslin. The role of embodied intention in early lexical acquisition. In *25th Annual Meeting of Cognitive Science Society (CogSci 2003)*, 2003.

Exploring Android Developmental Psychology in a Simulation World

Ben Goertzel (ben@goertzel.org)
Applied Research Lab for National and Homeland Security, Virginia Tech, 2000 N. 15t St., Ste. 50
Arlington, VA 22216 USA

Ari Heljakka (heljakka@iki.fi), Stephan Vladimir Bugaj (stephan@bugaj.com),
Cassio Pennachin (cassio@vettalabs.com)
Novamente LLC, 1405 Bernerd Place
Rockville, MD 20851 USA

Moshe Looks (moshe@metacog.org)
Department of Computer Science, Washington University, One Brookings Drive
St. Louis, MO 63130 USA

Abstract

A research programme oriented toward exploring the development of android cognition in the context of a 3D simulation world is described. The simulation world itself, AGISim, is an open-source project built on a game engine, which allows human-controlled and AI-controlled android agents to interact inside a simulated apartment. AGISim has been used for the embodiment of android agents controlled by the Novamente AI Engine, in the context of an AI-education program guided by Piagetian learning theory. Recent experiments have focused on teaching these android agents to understand the notion of the permanent identity of an agent or an object. Experiments involving the spontaneous learning of "theory of mind" have been designed and will be performed in the near future.

Introduction

The main stream of research in android science focuses, as it should, on the engineering, programming and instruction of physical android robots. However, the current state of android hardware is still relatively primitive, which means that in practical research, cognition tends to get short shrift, since achieving effective android perception and action are still such major obstacles. Thus, we suggest that it is worthwhile to complement work on physical android robotics with work in computer simulation worlds, in which human-controlled simulated androids interact with AI-controlled androids in simulated environments.

No simulation world running on currently affordable hardware will be able to provide a fully accurate simulation of the perceptual and motor-control challenges faced by physical android robots. However, we suggest that contemporary simulation worlds, appropriately utilized, can nonetheless permit effective simulation of many of the cognitive challenges that physical android robots face.

With this philosophy in mind, we have created a 3D simulation world called AGISim (Goertzel et al, 2006), and begun using it to teach an AI system to control a simulated android, in the context of interactions with a human-controlled simulated android. Within this framework we are pursuing an AI-teaching program loosely guided by Piagetian developmental psychology. Our current focus is on infant-level cognition such as the understanding of the permanence of objects and agents, exemplified for instance by Piaget's A-not-B task. The next phase of teaching will focus on "theory of mind" – on encouraging the AI system to come to its own understanding of the intentions and beliefs and knowledge of other cognitive agents, based on its interactions with them in the simulated world.

AGISim and Novamente

The AGISim simulated world is being developed as an open-source project[1], led by the first two authors, and is based on the CrystalSpace[2] 3D game engine, which may be configured to display realistic physics. It allows AI systems and humans to control android agents, and to experience the simulated world via multiple senses, as well as having the capability to chat with each other directly through text.

It is intended that the experience of an AGI controlling an agent in AGISim should display the main qualitative properties of a human controlling their body in the physical world. The simulated world should support the integration of perception, action and cognition in a unified learning loop. And, it should support the integration of information from a number of different senses, all reporting different aspects of a common world. With these goals in mind, we have created the initial version of AGISim as a basic 3D simulation of the interior of a building, with simulations of sight, sound, smell and taste. An agent in AGISim has a certain amount of energy, and can move around and pick up objects and build things. While not an exact simulation of any specific physical robot, the android agent an AI controls in AGISim is designed to bear sufficient resemblance to a simple humanoid robot that the porting of control routines learned in AGISim to a physical robot should be feasible, though certainly not trivial.

Our work with AGISim to date has focused on controlling android agents in AGISim using the Novamente AI Engine (or NAIE; Goertzel, 2006; Looks, Goertzel and Pennachin, 2004), a comprehensive AI architecture that synthesizes perception, action, abstract cognition, linguistic

[1] sourceforge.net/projects/agisim
[2] crystal.sourceforge.net

capability, short and long term memory and other aspects of intelligence, in a manner inspired by complex systems science. Its design is based on a common mathematical foundation spanning all these aspects, which draws on probability theory and algorithmic information theory among other areas. Unlike most contemporary AI projects, it is specifically oriented towards artificial *general* intelligence (AGI), rather than being restricted by design to one narrow domain or range of cognitive functions.

The NAIE integrates aspects of prior AI projects and approaches, including probabilistic inference, evolutionary programming and reinforcement learning. However, its overall architecture is unique, drawing on system-theoretic ideas regarding complex mental dynamics and associated emergent patterns. The existing code base implements roughly 60% of the overall design, and is being applied in bioinformatics, NLP and other domains.

Cognitive Development in Simulated Androids

Jean Piaget, in his classic studies of developmental psychology (Inhelder and Piaget, 1958), conceived of child development as falling into four stages, each roughly identified with an age group: infantile, preoperational, concrete operational, and formal. While Piaget's approach is out-of-date in some ways, recent researchers have still found it useful for structuring work in computational developmental psychology (Shultz, 2003). As will be described below in some detail, we have modified the Piagetian approach somewhat for usage in our own work.

The basic Piagetian stages are as follows:

- *Infantile*: Imitation, repetition, association. Object permanence – infants learn that objects persist even when not being observed.
- *Preoperational*: Abstract mental representations. Word-object and image-object associations become systematic rather than occasional. Simple syntax.
- *Concrete*: Abstract logical thought applied to the physical world: conservation laws; more sophisticated classification; theory of mind – an understanding of the distinction between what I know and what others know. Classification becomes subtler.
- *Formal*: Abstract deductive reasoning contextually and pragmatically applied, the process of forming then testing hypotheses, etc.

We have carried out learning experiments involving the NAIE and AGISim, corresponding to aspects of Piaget's early stages of development. We have shown that a Novamente-powered simulated android can learn, via its interactions with human-controlled simulated android agents, to carry out basic cognitive tasks like word-object association and understanding the permanence of objects and agents. A little later we will discuss a particular example related to object permanence: we have created a simulation of the "A-not-B" task commonly used in developmental psychology (see e.g. Thelen and Smith, 1994), and shown that the NAIE's ability to solve this task is specifically tied to its correct use of a specific inference rule called the "Rule of Choice."

We have also shown that an NAIE-powered simulated android can successfully classify objects in the simulation world, at a level corresponding roughly to Piaget's preoperational phase. A typical experiment involves distinguishing chairs from couches from boxes based on their appearance in various contexts.

Piagetian Stages and Uncertain Inference

To address the various critiques of Piaget's theories that have been made (Commons et al, 1982, 1988; Pascual-Leone and Smith, 1989; Fischer, 1980; Marchand, 2001), and to better bridge the gap between the human developmental psychology on which Piaget's theory was based and the non-human cognitive structures within the NAIE, we have created a novel theory of cognitive developmental stages, defined in terms of the control of uncertain inference trajectories. Our theory of developmental stages is oriented specifically toward AI systems like Novamente that are founded on uncertain logical inference rules "control schemata" (sometimes learned, sometimes pre-programmed) that determine the order in which inference rules are to be applied. It may be that our modified version of Piagetian theory also has applicability to human psychology, but at the moment we are focusing on its applications to AI and humanoid robotics.

The stages in our theory are defined as follows:

1. *Infantile*: Able to recognize patterns in and conduct inferences about the world, but only using simplistic hard-wired (not experientially learned) inference control schemata, along with pre-heuristic pattern mining of experiential data.
2. *Concrete:* Able to carry out more complex chains of reasoning regarding the world, via using inference control schemata that adapt behavior based on experience (reasoning about a given case in a manner similar to prior cases).
3. *Formal*: Able to carry out arbitrarily complex inferences (constrained only by computational resources) via including inference control as an explicit subject of abstract learning.
4. *Reflexive*: Capable of self-modification of internal structures. (In the case of the NAIE, this process is very direct and thorough: one key architectural difference between humans and AI's is the latter's vastly greater capability for self-modification)

In this approach Piaget's preoperational phase appears as transitional between the infantile and concrete operational phases; and, following a number of recent thinkers, we have explicitly introduced a post-formal stage as well.

The semantics of our stages is similar but not identical to Piaget's. Our stages are defined via internal cognitive mechanisms, which represent not only abstract knowledge, but also perceptual information, and both cognitive and

operational task skills. Uncertain inference is assumed to allow the developing cognitive system to reason through experience in a fuzzy and context-variant way, rather than requiring a fully-formed absolute logical formulation of each particular situation.

We posit that these developmental stages correspond to the ability to solve certain classes of problems in a generalizable way. For instance, we suggest that it is only through inference control schemata which adapt based on experience that uncertain inference-based AI systems can learn to consistently solve Piagetian concrete-operational tasks in a way that provides knowledge suitable for further generalization. Of course, it may be that minds using hard-wired inference control schemata (typical of the infantile stage) can still solve some Piagetian concrete-operational tasks. Such brittle approaches to solving such tasks, historically, have proved unable to generalize sufficiently and resulted in permanently "brittle" AI systems which are very limited in capability.

We have designed an AGISim based learning programmed for the NAIE based on these stages, while at the same time accounting for key ideas of dynamical developmental psychology:

1. Not all tasks will develop into the same stage at the same time.
2. Stages represent not only abstract cognitive abilities, but also interrelated perceptual and operational abilities.

Since development is a dynamical feedback system between perception, action, and abstraction, different abstract cognitive tasks may develop into different stages at different times based on the amount of experience with relevant actions and perceptions.

The A-not-B Error

In Piaget's classic "A-not-B error," the teacher hides an object in location A repeatedly, then eventually hides it in location B and asks the subject (a baby or in our case, an AI agent) to find it (Thelen and Smith, 1994; see also Spencer et al, 2001).. Human babies less than 9 months of age who have successfully uncovered a toy at location A in prior trials will often continue to reach to that location even after they watch the toy hidden in a nearby location B. Older babies will look in B after they've seen the toy hidden in B. In some cases, infants will look at B (the correct location) but reach for A (the incorrect location), indicating a complex and only semi-consistent internal knowledge base.

We have created a simulated-humanoid-robotics emulation of this Piagetian scenario in AGISim, and used the NAIE in this context. The simulation involves a humanoid agent controlled by the NAIE, and a humanoid teacher agent controlled by a human. The A-not-B task is represented via the placement of a toy bunny in boxes which the agents may open or close by pushing buttons. Thus, rather than an abstract mathematical problem, the NAIE-controlled agent is presented with a problem involving integrated perception, action and cognition, analogous to the problems presented by human infants.

What we find is that, like older babies, the NAIE learns through interactive experience to look in location B – it learns that objects exist even when unobserved. However, it is also possible to emulate younger-baby behavior within the NAIE by modifying the way its inference rules operate.

The NAIE system's ability to solve the A-not-B task correctly is specifically tied to its correct use of a specific inference rule called the "Rule of Choice." This is the rule that allows the system's inference engine to correctly choose between two competing estimates of the truth value of the same relationship. In this case, one estimate comes from the simple repetitive fact that the toy has often been found in location A before; and another estimate comes from the knowledge that inanimate objects tend to be found where they were last seen. The latter knowledge is more general and a properly functioning Rule of Choice will choose it over the former in this case.

We can also emulate, in the NAIE, the infantile behavior in which the A-not-B error is made via reaching but not via looking. This occurs when implications joining sensations to actions are made and revised directly without the intervention of more abstract representations that abstract away from particular sensation and action modalities. This may happen in the NAIE when the parameters of the concept-formation and inference components are not tuned to encourage the learning of abstractions.

The inferences involved in the A-not-B task are "infantile" in the sense of our above developmental theory, in the sense that they can be carried out using simple non-adaptive forward-chaining or backward-chaining inference control. The difference between incorrect and correct behavior on this task has to do, not with the high-level properties of inference control schemata, but with the correct usage of a particular inference rule. However, it is the execution of specific inferences like this one from which general inference control patterns are learned, enabling adaptive inference control as is required by later developmental stages.

Conclusion

We have described a research programmed aimed at teaching an AI system to control a humanoid agent in a simulated environment, and discussed some of the early steps we have taken along this path.

The next stage of our work will involve Piaget's concrete operational phase, and in particular "theory of mind." We aim to teach a Novamente-controlled simulated android agent to understand that other simulated android agents possess knowledge and beliefs separate from their own. It is important to stress that in our approach this knowledge is not programmed: it is learned based on embodied social interaction.

To teach "theory of mind," for example, the teacher can hide an object in such a way that Novamente can see that it sees the hiding action but its playmate cannot; and then it can be checked if Novamente can predict where the playmate will look for the object. Without adequate theory of mind, Novamente will predict the playmate will look in the place the object was hidden. With adequate theory of mind, Novamente will predict the playmate will search for the object in the most obvious place.

The knowledge obtained via this sort of experiment is not particular to the AGISim simulation world, but is of generic value, and should be largely portable to the domain of physical android robotics, at such time when the NAIE is used to control a physical android robot. And the lessons learned in doing this work should be in large measure extensible beyond the NAIE to AI-driven humanoid robotics in general.

References

Commons, M., F. Richards & D. Kuhn. (1982). Systematic and metasystematic reasoning: a case for a level of reasoning beyond Piaget's formal operations. *Child Development, 53*, 1058-1069.

Commons, M., Trudeau, E.J., Stein, S.A., Richards, F.A., & Krause, S.R. (1998). Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review, 18*, 237-278.

Fischer, K. (1980). A theory of cognitive development: control and construction of hierarchies of skills. *Psychological Review. 87*, 477-531.

Goertzel, B. (2006). Patterns, hypergraphs and artificial general intelligence. *Proceedings of IJCNN 2006, Vancouver CA.*

Goertzel, B., M. Looks, A. Heljakka, & C. Pennachin. (2006). Toward a pragmatic understanding of the cognitive underpinnings of symbol grounding, in *Semiotics and Intelligent Systems Development*, Edited by Ricardo Gudwin and João Queiroz.

Inhelder, B. and J. Piaget. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Looks, M., B. Goertzel, & C. Pennachin. (2004). Novamente: an integrative architecture for Artificial General Intelligence. *Proceedings of AAAI 2004 Symposium on Achieving Human-Level AI via Integrated Systems and Research*, Washington, DC.

Marchand, H. (2001). Reflections on PostFormal thought. In *The Genetic Epistemologist, 29*(3).

Pascual-Leone, J. and J. Smith. (1969). The encoding and decoding of symbols by children: a new experimental paradigm and neo-Piagetian model. *Journal of Experimental Child Psychology, 8*, 328-355.

Shultz, T. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.

Thelen, E. & L. Smith (1994). A dynamic systems approach to the development of cognition and action. Cambridge, MA: MIT Press.

Spencer, J.P., L.B. Smith, and E. Thelen. (2001). Biobehavioral development, perception, and action tests of a dynamic systems account of the A-not-B error: The influence of prior experience on the spatial memory abilities of two-year-olds. *Child Development*, vol. 72 issue 5, Sept/Oct 2001, 1327.

Thelen, E., G. Schoner, C. Scheier, and L.B. Smith. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral Brain Science*, 24(1), 1-34.

Disappearance of Inversion Effect for Walking Animation with Robotic Appearance

Masahiro Hirai (hirai@ardbeg.c.u-tokyo.ac.jp)
Department of General System Studies,
The University of Tokyo,
3-4-1 Komaba, Meguro-ku, Tokyo, 153-8902 JAPAN

Kazuo Hiraki (khiraki@idea.c.u-tokyo.ac.jp)
Department of General System Studies,
The University of Tokyo,
3-4-1 Komaba, Meguro-ku, Tokyo, 153-8902 JAPAN

Abstract

Recent studies have reported similarity in the neural processing of human and robot actions; however, whether this is the case remains controversial. Here, we examined this controversy using the inversion effect, a phenomenon whereby an upright face- and body-sensitive event-related potential component is enhanced and delayed in response to an inverted face and body, but not an inverted object. The results showed that the inversion effect occurs only with a human, not with robotic and point-light appearances, suggesting that our visual system differentially processes human and robot actions.

Introduction

It has been suggested that our neural system is tuned specifically to be able to detect the human body. For example, a previous psychophysical study revealed that the neural system differentially processes the human body and objects (Shiffrar & Freyd, 1990), while recent neuroimaging studies have shown specific tuning to the human body (Downing et al., 2001), the human face (Gauthier et al., 2000; Kanwisher, 2000) and human body movements (e.g. biological motion; Grossman et al., 2000). These findings imply that our neural system responds sensitively to both human appearance and motion.

With the recent development of robotic technologies, living with robots has become a reality, not just something seen in science fiction movies. Moreover, various kinds of robots now appear in our daily lives; for example, humanlike robots (Collins et al., 2005) such as "ASIMO" [1], "QRIO"[2] and "Robovie" (Ishiguro et al., 2003), which were designed and developed specifically for household use. Furthermore, robots with a very human-like appearance are now being developed, and at a glance, are often indistinguishable from human beings [3].

Since these robots have similar appearance information to humans, such as body structure and configuration, yet are not a biological object, the question therefore arises as to whether or not our neural system interprets such robots as a kind of human. To date, several studies have provided clues to answer this question. For example, in a behavioral study, Kilner et al. (2003) reported that observation of other humans making incongruent movements, but not robots, had a significant interference effect on executed movements of participants. On the other hand, Pelphrey et al. (2003) reported that activation of the superior temporal sulcus (STS) during processing of the human appearance is similar to that during processing of a robotic appearance. The former study suggests that superficial information might affect our perception-action system, while the latter suggests that motion information, not just superficial information, might also affect activation of the STS.

Intuitively, both appearance and motion information therefore seem to play an important role in detecting characteristics of 'human-likeness'. That is, our visual system discriminates humans from objects not only by detecting appearance information, but also using motion information such as biological motion perception (Johansson, 1973). However, the relationship between appearance and motion information in detecting 'human-likeness' has not been fully investigated.

The aim of the present study is to clarify how different appearance information with identical motion information affects the neural response. To investigate this, we recorded event-related potentials (ERPs) in human participants and evaluated the occurrence of the inversion effect (Bentin et al., 1996; Linkenkaer-Hansen et al., 1998; Rossion et al., 2000; Taylor et al., 2001; Itier & Taylor, 2004; Stekelenburg & de Gelder, 2004). The inversion effect is a phenomenon whereby an upright face- and body-sensitive ERP component (N170) is delayed (Bentin et al., 1996; Linkenkaer-Hansen et al., 1998; Taylor et al., 2001) and enlarged in amplitude (Linkenkaer-Hansen et al., 1998; Rossion et al., 2000; Taylor et al., 2001; Itier & Taylor, 2004) in response to inverted faces and bodies but not inverted objects (Rossion et al., 2000; Stekelenburg & de Gelder, 2004). An inversion effect has also been reported in magnetoencephalography (MEG) (Watanabe et al., 2003) and functional magnetic resonance imaging (fMRI) (Haxby et al., 1999) studies of upright and inverted face perception.

In this study, we employed three kinds of walking animation with different superficial information (human, robot and point-light appearance) to explore two hypotheses. The first hypothesis is that if robotic walking animation is processed like an object, the inversion effect will not occur. However, in contrast, if it is processed like human information, the inversion effect will be observed as in the human appearance condition. The second hypothesis

[1] http://asimo.honda.com/
[2] http://www.sony.net/SonyInfo/QRIO/
[3] http://news.bbc.co.uk/1/hi/sci/tech/4714135.stm

is that if superficial information does not affect processing of human walking animation, ERP waveforms in all three conditions will show similar patterns because of the identical walking actions. Both hypotheses were tested by measuring ERPs.

Materials and Methods

Three kinds of walking animation (Fig. 1) (human, robot and point-light) with two orientations (upright and inverted) were employed. The structure of the body and walking speed were identical in all animations. Nineteen healthy participants were included as study participants. They were required to view each animation passively and mentally count the number of asterisks appearing randomly during each block. Electroencephalograms (EEGs) were recorded during each trial with a Geodesic Sensor Net composed of 64 electrodes (Tucker, 1993).

Participants

We studied nineteen healthy participants (range/mean age: 18-30/23.7 ± 3.9 years; 14 males, 5 females). Seventeen subjects were right-handed and all had normal or corrected-to-normal vision. All subjects provided informed consent for a protocol that was approved by the Ethics Committee of the University of Tokyo.

Experimental Procedure

Six experimental conditions were employed as shown in Fig.1. To generate the animated figures, we used the Poser 5.0 software program (Curious Labs, Santa Cruz, CA). Both the human and robot animations were generated using built-in 3D models. For the point-light animation, the human 3D model used in the human animation was replaced by 14 small balls placed at all joints and the head using Metasequoia (Mizuno Lab, Japan).

All animations were viewed in profile as walking as if on a treadmill. The walking speed in all animations was 2.0 steps per second. The animations were displayed on a 17-inch monitor against a black background. Each participant was seated 100 cm from the display in a dimly lit room. The entire visual stimulus was approximately 3 × 3°. To produce smooth animated motion, each animation comprised 15 frames displayed for 510 ms and with an interframe interval of approximately 34 ms. The initial number of frames was randomized to prevent the participants from remembering the initial starting figure.

Each experiment consisted of eight blocks with a 1 min inter-block interval. Twenty stimuli were employed in each animation condition, and accordingly, 120 animations were presented per block and 960 per experiment. Thus, each animation was presented 160 times throughout each experiment. In each trial, the stimulus was presented for 510 ms followed by presentation of a white fixation point (a 0.3 × 0.3° cross) for 500 ms. To ensure that subjects maintained their gaze on the center of the monitor during all animations, participants were asked to engage in a continuous performance task. They were asked to count the number of times a yellow asterisk appeared randomly on the screen and report back at the end of each block. The asterisk was presented for 500ms instead of the walking animation eight times per block.

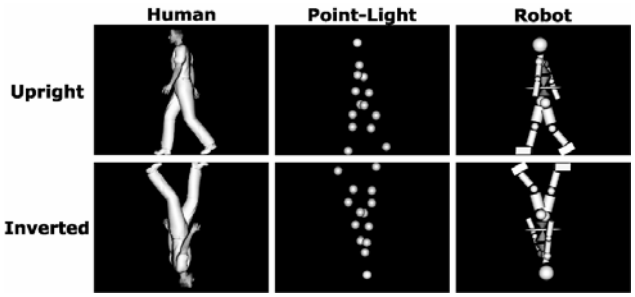

Figure 1: The experimental stimuli. All animations were viewed in profile as walking as if on a treadmill.

Results

Behavior results

The percentage of correct performance in the counting task was 46.1 ± 24.3 % (average ± S.D).

ERP results

Fig.2 shows the grand mean waveforms of the ERP responses. As in our previous study (Hirai et al., 2005), we collapsed the three electrodes surrounding each T5/T6 (International 10-20 System) into two sites. A single negative peak was found at around 200 ms (conventional N170-like component) in both the human and robot conditions, while in the point-light condition two negative peaks were observed at 200 and 340 ms, respectively. The peak latency and amplitude (in order to correct the N1 amplitude, we calculated the P1-N1 amplitude) of each component were also calculated, and subsequently, statistical analysis was carried out.

P1-N1 amplitude and N1 latency In three-way ANOVA of the P1-N1 amplitude, laterality × type of appearance × orientation was significant [$F(2,36) = 3.37$, $p<0.05$]. Subsequent analysis revealed that the amplitude in the right hemisphere was significantly larger with the inverted orientation than the upright orientation in the human appearance condition [4.24μV vs. 5.18μV, $F(1,108) = 9.62$, $p<0.01$]. In addition, the amplitude in the left hemisphere was significantly larger than that in the right hemisphere in the upright-human condition [5.09 μ V vs. 4.24μV, $F(1,108) = 4.14$, $p<0.05$]. Moreover, the amplitude with the human condition was significantly larger than that with the point-light condition in the left hemisphere with the upright orientation [5.09μV vs. 3.81μV, $p<0.01$; Tukey's HSD]. The amplitude with the human condition was also significantly larger than that with the point-light condition in both hemispheres with the inverted orientation [left hemisphere: 5.29μV vs. 4.24μV, $p<0.05$; right hemisphere: 5.18μV vs. 3.89μV, $p<0.01$; Tukey's HSD]. Similarly, the amplitude with the human condition was significantly larger than that with the robot condition in both hemispheres with the inverted orientation [left hemisphere: 5.29μV vs. 4.39μV,

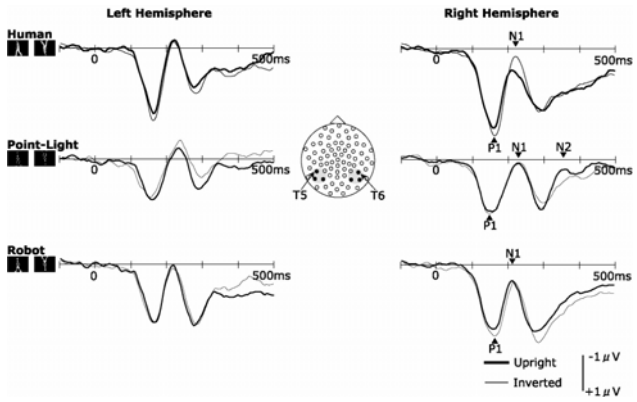

Figure 2: Grand averaged ERP waveforms showing the inversion effect with each appearance and orientation condition.

$p<0.05$; right hemisphere: 5.18μV vs. 4.13μV, $p<0.05$; Tukey's HSD]. The main effect of orientation was also significant with the N1 latency, [$F(1,18) = 4.72$, $p<0.05$; upright: 218.6ms vs. inverted: 224.0ms], indicating that the latency of inverted stimuli was longer than that of upright stimuli.

P1-N2 amplitude and N2 latency The N2 component was observed only with the point-light condition, and accordingly, the P1-N2 amplitude and N2 latency were analyzed as above. Two-way of ANOVA was applied to the P1-N2 amplitude using laterality (left or right) and orientation (upright or inverted) as variables. As a result, laterality × orientation was shown to be significant [$F(1,18) = 10.9$, $p<0.01$]. Subsequent analysis revealed that the amplitude with the inverted condition was significantly larger than that with the upright condition in the left hemisphere [4.03μV vs. 3.48μV, $F(1,36) = 6.28$, $p<0.01$]. On the contrary, the amplitude with the upright condition was significantly larger than that with the inverted condition in the right hemisphere [3.68μV vs. 3.01μV, $F(1,36) = 9.74$, $p<0.01$]. Moreover, the amplitude in the left hemisphere was significantly larger than that in the right hemisphere with the inverted condition [4.03μV vs. 3.01μV, $F(1,36)=5.53$, $p<0.01$]. No significance was observed regarding the N2 latency in the point-light motion condition.

The P1-N1 amplitude in the right hemisphere was significantly larger with the inverted orientation than the upright orientation with the human appearance condition [4.24μV vs. 5.18μV, $F(1,108) = 9.62$, $p<0.01$]. This was not observed with the other appearance conditions. Regarding the N1 latency, the main effect of orientation was also significant [$F(1,18) = 4.72$, $p<0.05$; upright: 218.6ms vs. inverted: 224.0ms], indicating that the latency of the inverted stimuli was longer than that of the upright stimulus.

Conclusion and Discussion

Our data demonstrated that the inversion effect occurs in the right occipitotemporal region with the human appearance condition only. These findings are consistent with the results of recent neuroimaging studies of face and body perception (Bentin et al., 1996; Linkenkaer-Hansen et al., 1998; Rossion et al., 2000; Taylor et al., 2001; Itier & Taylor, 2004; Stekelenburg & de Gelder, 2004). With regard to latency, a recent study suggested that the delay in latency of the N1 component is observed not only with faces but also objects (Itier et al., 2006), which is also consistent with our present data. The present findings imply that robot walking animation is not processed like human information (i.e. robots are not categorized as humans), even though the robots are analogous in appearance and have identical motion properties (speed and motion trajectory). This suggests that appearance information affects the neural responses and this categorization is processed within early visual processing.

In the light of recent findings, our data seems inconsistent with those of Pelphrey et al. (2003) who suggests the importance of motion information only. That is, our results show that appearance information also has an affect on the neural response in the occipitotemporal region, which might be involved with the STS region (Homan et al., 1987). However, for the following two reasons, we believe these findings are in fact consistent. First, Pelphrey et al. (2003) used an fMRI technique to investigate conditional differences, and thus, could not measure the neural response with millisecond temporal resolution. On the other hand, in our study, the conditional differences were observed at around 200ms after stimulus onset, a rapid response that neuroimaging techniques such as fMRI are perhaps unable to detect. The second reason is related to their use of upright and scrambled conditions only; that is, they did not investigate the inversion effect. Consistent with their fMRI findings, our ERP data showed no conditional difference in the P1-N1 amplitude between the upright-robotic and upright-human conditions.

As in our previous study, we found two negative peaks in the occipitotemporal region at around 200 and 340ms, respectively, with the point-light motion condition. The second negative component is thought to reflect processing of biological motion (Hirai et al., 2003, 2005), specific analysis of motion patterns providing biologically relevant information (Jokisch et al., 2005) or form-from-motion processing (Wang et al., 1999).

The conditional differences in the inversion effect in the present study might also be explained from the point of view of perceptual expertise. Several studies have reported that for such experts (e.g. dog show judges) processing of a car, dog, or bird is similar to processing of a face (Gauthier & Tarr, 2002; Diamond & Carey, 1986). Another ERP study also showed inversion of the N170 component in response to a non-face object (Greebles) as well as faces with expertise training in Greebles (Rossion et al., 2002). Accordingly, it is likely that an object with a robotic appearance is not observed frequently, unlike the human body, and thus, this frequency of contact might have elicited the conditional difference in the inversion effect.

In conclusion, using the inversion effect as an index, this study clarified that a human walking appearance is processed differently from robotic and point-light walking appearances. These findings indicate that our visual system distinctly and differentially processes biological

appearances such as the human body and nonbiological appearances such as robots within a short latency, even when the motion information is superimposed. To determine the role of motion information, further work is needed to fully elucidate the differential neural responses to robotic and human motion with identical appearance information.

Acknowledgments

This work was supported by Grants-in-Aid from MEXT, Japan (#15017214), and the 21st century COE program (Center for Evolutionary Cognitive Sciences at the University of Tokyo), Japan.

References

Bentin, S., Allison, T., Puce, A., Perez, A., McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience,* 8, 551-565.

Collins, S., Ruina, A., Tedrake, R., Wisse, M. (2005). Efficient bipedal robots based on passive-dynamic walkers. *Science,* 307, 1082-1085.

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

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

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

Gauthier, I., & Tarr, M.J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance,* 28, 431-446.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan V., Neighbor, G., Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, 12, 711-720.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L., Hoffman, E.A., Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron,* 22, 189-199.

Hirai, M., Fukushima, H., Hiraki, K. (2003). An event-related potentials study of biological motion perception in humans. *Neuroscience Letters,* 344, 41-44.

Hirai, M., Senju, A., Fukushima, H., Hiraki, K. (2005). Active processing of biological motion perception: an ERP study. *Brain Research: Cognitive Brain Research,* 23, 387-396.

Homan, R.W., Herman, J., Purdy, P. (1987). Cerebral location of international 10-20 system electrode placement. *Electroencephalography and Clinical Neurophysiology*, 66, 376-382.

Ishiguro, H., Ono, T., Imai, M., Kanda, T. (2003). Development of an interactive humanoid robot "Robovie" -An interdisciplinary approach, R. A. Jarvis and A. Zelinsky (Eds.), *Robotics Research,* Springer, pp. 179-191.

Kilner, J.M., Paulignan, Y., Blakemore, S.J. (2003). An interference effect of observed biological movement on action. *Current Biology,* 13, 522-525.

Itier, R.J., & Taylor, M.J. (2004). Source analysis of the N170 to faces and objects. *Neuroreport*, 15, 1261-1265.

Itier, R.J., Latinus, M., Taylor, M.J. (2006). Face, eye and object early processing: what is the face specificity? *Neuroimage,* 29, 667-676.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics,* 14, 201-211.

Jokisch, D., Daum, I., Suchan, B., Troje, N.F. (2005). Structural encoding and recognition of biological motion: evidence from event-related potentials and source analysis. *Behavioral Brain Research,* 157, 195-204.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3, 759-763.

Linkenkaer-Hansen, K., Palva, J.M., Sams, M., Hietanen, J.K., Aronen, H.J., Ilmoniemi, R.J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters,* 253, 147-150.

Pelphrey, K.A., Mitchell, T.V., McKeown, M.J., Goldstein, J., Allison, T., McCarthy, G. (2003). Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *Journal of Neuroscience*, 23, 6819-6825.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M.J., Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science,* 13, 250-257.

Rossion, B., Gauthier, I., Tarr, M.J., Despland, P., Bruyer, R., Linotte, S., Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, 11, 69-74.

Shiffrar, M., & Freyd, J.J. (1990). Apparent motion of the human body. *Psychological Science,* 1, 257-264.

Stekelenburg, J.J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. *Neuroreport*, 15, 777-780.

Taylor, M.J., Edmonds, G.E., McCarthy, G., Allison T. (2001). Eyes first! Eye processing develops before face processing in children. *Neuroreport,* 12, 1671-1676.

Tucker, D.M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, 87, 154-163.

Wang, J., Jin, Y., Xiao, F., Fan, S., Chen, L. (1999). Attention-Sensitive Event-Related Potentials Elicited by Kinetic Forms. *Clinical Neurophysiology*, 110, 329-341.

Watanabe, S., Kakigi, R., Puce, (2003). A. The spatiotemporal dynamics of the face inversion effect: a magneto- and electro-encephalographic study. *Neuroscience,* 116, 879-895.

Mentalizing to Non-human Agents by Children

Shoji Itakura (sitakura@bun.kyoto-u.ac.jp)
Kyoto University & ATR Intelligent Robotics and Communication Lab.

Takayuki Kanda (kanda@atr.jp)
ATR Intelligent Robotics and Communication Lab.

Hiroshi Ishiguro (ishiguro@ams.eng.osaka-u.ac.jp)
Osaka University & ATR Intelligent Robotics and Communication Lab.

Abstract

In Experiment 1, fifty two-year-old children were tested to examine whether they could reproduce the target outcome of a robot in a goal re-enactment paradigm developed by Meltzoff (1995). The results show that the children were not only able to reproduce the target action produced by the robot, but were also able to complete the same task which the robot attempted, but failed to perform. However, it was essential that the robot mimicked human behavior suggesting intention, such as gazing at a partner and at the object being manipulated, in order to induce children to produce the target outcome in the failed attempt condition. In Experiment 2, a standard False Belief Task was conducted with a robot to investigate whether preschoolers attribute false belief to a robot or not. Results suggested that the children attribute false belief to a robot but don't attribute a mental verb to it.

Introduction

When do children first attribute mental states to others, and when they do, to whom do they attribute the mental state? Several studies have suggested that children comprehend goal-directed behavior from an early age (Carpenter et al., 1998; Csibra, 2003; Frith & Frith, 2003; Gergely et al., 1995). Woodward (1998) developed a new paradigm for understanding goal-directedness using visual habituation. She tested whether infants encode human action in terms of an actor's goals or in terms of spatiotemporal movement. In her experiments, infants viewed a hand reaching towards one of two objects. Upon habituation, the location of these objects was switched and the experimenter reached either towards the other object in the same location or the original object in its new location. It was observed that both five- and nine-month-old infants looked for longer when the experimenter reached towards the new object than when she grasped the old object in its new position. Woodward therefore concluded that young children tend to encode the actions of other people as goal-directed. These results, and those of subsequent studies, suggest that infants attribute an intentional relationship between objects and the world (Johnson, 2000). Meltzoff (1995) produced further evidence of goal comprehension in infants using the re-enactment of goals paradigm. In his study, 18-month-old infants reproduced the aims of the object-directed actions of adults, even in cases when the goals set within the model were never actually attained, but had to be inferred. However, under conditions in which the human agent was replaced by mechanical pincers performing the same actions, infants did not achieve the unattained goals. Meltzoff concluded that whereas 18-month-old infants were able to gauge the intentions of a human and complete the failed action, this was not the case with a mechanical pincer, to which they did not attribute goals. Johnson, Booth, and O'Hearn (2001) studied infant imitation and the production of communicative gestures, starting from the hypothesis that the recognition of mentalistic agents is not isomorphic with person recognition. Rather, it is based on non-arbitrary object perception, including the presence of a face and the ability to interact contingently with other agents. These authors replicated Meltzoff's study but modified it by using a stuffed orangutan as the non-human agent. They found that 15-month-old infants re-enacted the goals of an inanimate object that had a face and interacted contingently with the infants and the experimenter. In Experiment 1, adopting the same perspective as that of Johnson, Booth, and O'Hearn (2001), and using Meltzoff's (1995) re-enactment of goals paradigm, we investigated firstly whether young children imitate the actions of an autonomous humanoid robot, and secondly, whether they attain the goals indicated by its incomplete action.

One of the most important milestones in social cognitive development is to understand another's false belief. The False Belief Task developed by Wimmer and Perner (1983), also known as the "Maxi Task," measured this ability. It goes as follows: Maxi has some chocolate and puts it into a blue cupboard. Then he goes out. Now his mother comes in and moves the chocolate to a green cupboard. Maxi comes back to get his chocolate. Where will Maxi look for the chocolate? The answer is Maxi will look in the blue cupboard, because this is where he erroneously believes the chocolate to be. A series of studies established that children of around four years old begin to pass this task and can verbally explain it when asked.

In Experiment 2, we conducted the False Belief Task with a robot to establish whether preschoolers attribute false belief to a robot or not.

Experiment 1: Inference of a robot's goal by young children

Method

Participants: The sample consisted of 50 children (24 boys and 26 girls). Seven additional infants were excluded because they failed to complete all the test trials. Participants were all aged between 24 and 35 months (M=30.6, SD=3.2).

Stimuli: The experiment took place in an infant laboratory at Kyoto University. We employed an autonomous robot named Robovie, developed at the ATR Intelligence Robotics Laboratory in Kyoto, Japan. Robovie is an autonomous humanoid robot (1.2 m tall, with a 50 cm radius, and weighing 40 kg) that can move independently, and has human-like eyes and hands. It is equipped with visual, auditory, and tactile sensors, designed to enable it to imitate human behavior. Robovie can engage in communicative behavior with humans and mimics human behavior such as shaking hands, joint visual attention, and pointing.

In the present study, unlike the experiments of Meltzoff (1995) and Johnson, Booth, and O'Hearn (2001), the agent's action was presented to the children on a video monitor (38 x 64 cm). We considered it reasonable to use a video monitor to present these stimuli, as Barr and Hayne (2000) previously reported that 18-month-old infants imitated target actions in a video monitor condition. Phillipes and Wellman (in press) have also demonstrated the validity of using videotaped actions in research with infants. There were two action trials, a full-demonstration (complete) action, and a failed-attempt (incomplete) action.

Each action (complete, incomplete) trial consisted of two gaze conditions according to the robot's gazing behavior in relation to a human social partner (Figure 1). Thus there were four types of video stimuli and a baseline condition. Each video stimulus lasted 20 seconds, and included the successive manipulation of three different sets of objects.

1) Full Demonstration + Gaze: The infant watched the robot act on each set of objects three times successfully. The robot gazed at its partner's face before beginning each task, then looked at the object while manipulating it, and finally gazed at its partner's face again after completing each set of actions.

2) Full Demonstration + No Gaze: The subject watched the robot act on each set of objects three times successfully, but unlike in the gaze condition, the robot kept looking forward during the task.

3) Failed Attempt + Gaze: The subject watched the robot act unsuccessfully on each set of objects three times. The robot gazed at its partner's face before beginning each task, then looked at the object while manipulating it, and finally gazed at the partner's face again after failing to complete each task.

4) Failed Attempt + No Gaze: The subject watched the robot act on each set of objects three times unsuccessfully, but unlike in the gaze condition, the robot kept looking forward during the task.

5) Baseline: In the baseline condition, each trial consisted of the child manipulating the object for 20 seconds without visual stimuli.

There were three sets of objects: a dumbbell, a cup and beads, and a peg with an elastic band.

The dumbbell. In the complete condition, the experimenter handed the object to the robot, which grasped one end of the dumbbell in each hand, pulling the two ends apart. For the incomplete condition, the robot grasped the dumbbell in the same manner, but one hand slipped off the end of the dumbbell before it came apart.

The cup and beads. In the complete condition, the experimenter handed the beads to the robot with the string above the edge of the cup and the robot subsequently dropped the beads inside the cup. For the incomplete condition, the robot grasped the beads, lifted the string above the edge of the cup, wavered slightly over it, and then dropped the beads outside the cup.

The peg and elastic band. In the complete condition, the experimenter handed the robot an elastic band, which it grasped and hung on the peg. For the incomplete condition, the robot grasped the elastic band, raised it up towards the peg, but released it just before it circled the peg, thus dropping it onto the table.

Procedure: During each session, the child was seated in front of a small table facing a video monitor, with his/her parent or caregiver seated behind or next to them. After a five-minute habituation period, the experimenter began operating the video monitor for the presentation of the stimuli. The upper half of the infant's body was monitored by a video camera placed under the video monitor.

After they had viewed the video stimuli, the object they had just seen the robot manipulate was placed in front of the child by the experimenter. The sequence of the three objects was fixed, as the order of presentation was not found to have a significant effect in previous studies (Meltzoff, 1995; Johnson et al., 2001). If the child did not touch the objects, the experimenter would call its name or say "Look!" to engage his/her attention, but did not give any direct instructions. The experimenter gave neither affective nor linguistic cues during the viewing of the video stimulus and the response period.

Results and discussion

Since there were three target actions, the score achieved by each infant ranged from 0–3. A child obtained 3 points if he/she completed the target action with all three object sets, and if he/she failed to complete the goal using any of the sets his/her score was 0. The mean score for the *Full Demonstration + Gaze* condition was 2.2; 2.1 for the *Full Demonstration + No Gaze* condition; 1.6 for the *Failed Attempt + Gaze* condition; 0.7 for the *Failed Attempt + No Gaze* condition, and 0.4 for the *Baseline Condition*. The resulting overall mean value for each condition is shown in Figure 2. An analysis of variance (ANOVA) was performed

on the effect of the *Gaze* condition on the *Full Demonstration*/*Failed Attempt* condition. A significant effect of the *Gaze* condition [$F(1, 36) = 4.29$, $P < 0.05$] on the *Full Demonstration*/*Failed Attempt* condition [$F(1, 36) = 17.14$, $P < 0.001$] was found. The interaction between the *Gaze* condition and the *Full Demonstration/Failed Attempt* condition was not significant [$F(1,36)=2.74$, $p <0.106$]. Only the *Failed Attempt + No Gaze* condition was not significantly different from the baseline condition [$t(18)=0.878$, n.s.].

No difference was observed in the children's performance, irrespective of whether or not the robot gazed at its partner's face during the full demonstration; the children imitated the robot's actions. The failed attempt, in which the children observed the robot's attempt and failure to produce the target outcome, was the most interesting. In this condition, the children produced the target outcomes when the robot looked at the partner and the object; however, they failed to produce the intended action when it did not exhibit such intention-implying behavior. In the baseline condition, the children rarely produced the target outcomes; this result is consistent with those obtained by Meltzoff (1995) and Johnson, Booth, and O'Hearn (2001).

Infants are known to distinguish between humans and inanimate objects. By two months of age, children treat people as social entities, smiling, vocalizing, and imitating their actions, but objects are treated as toys to be looked at and to be manipulated (Legerstee, 1991, 2001; Poulin-Dubois et al., 1996; Poulin-Dobois, 1999). Meltzoff (1995) also claimed that infants restrict their mental state attributions to people. In his study, when a human agent was replaced by a set of mechanical pincers, children failed to reproduce the incomplete action (Meltzoff, 1995: Experiment 2). However, Johnson, Booth, and O'Hearn (2001) replicated this experiment using a stuffed orangutan, demonstrating that a nonhuman agent could elicit the re-enactment of goal-orientated behavior by an infant under certain circumstances. The authors concluded that the agent needed to possess the features thought to characterize mentalistic agents, such as the ability to interact contingently with others, or the presence of a face (Johnson, Booth, and O'Hearn, 2001).

In the present study, when young children saw the robot "try" but fail to achieve the same set of target outcomes, and were given the objects they had seen the robot manipulate, they produced the inferred outcome, rather than the actually viewed event only when the robot showed intention-like actions, such as gazing. Following Meltzoff (1995), these responses were interpreted as evidence that the infant attributed goals to the agent. However, the children who saw the robot that did not gaze at its partner or the object in the same action condition (the incomplete condition) did not produce the target outcomes. This contrasts with the results of Meltzoff's (1995) study, in which the human demonstrator was not required to exhibit behavior implying an intention, such as gazing at the objects, in order to induce infants to produce the target outcome. These variances could be based on the ability of children to distinguish between humans and nonhumans.

Experiment 2: False belief task with a robot by preschoolers

Method

Participants: The participants were 58 young children (27 boys, 31 girls; range=54 months to 80 months; mean=65.4 months). We chose children of these ages because many studies demonstrated that children between the ages of four and five years start to pass the False Belief Task.

Materials: All the stimuli were presented on a video monitor. There were two versions of video stimuli. One of the video scenes was as follows: Robovie (see Experiment 1) puts the doll away in a particular location (Box A), then leaves the room. During Robovie's absence, the man removes the doll from Box A, and places the doll not back in Box A, but in Box B. The other video scene was the same as the robot version, except that a human was projected, instead of the robot.

Stimuli: Each subject was shown these two types of scenes, and given four questions just after watching each video scene individually. The order of presenting the stimulus was counterbalanced. Four questions are as follows: i) "Where will it/he look for a doll?" (Question for prediction); ii) "Which box does it/he think the doll is in?" (Question for representation); iii) "Which box contains a doll?" (Question for reality); iv) Which box contained a doll at the beginning of the session?" (Question for memory).

Results and discussion

The results are shown in Fig. 2. There was no difference between the human condition and the robot condition in the reality question ($z=0.01$, $P>0.992$, n. s.) and the memory question ($z=0.28$, $p>o.339$, n. s.). Most of the children answered these questions correctly. There was also no difference in the prediction question ($z=0.28$, $p>0.339$, n. s.) between both conditions. However, there was a significant difference between the human condition and the robot condition in the representation question ($z=3.68$, $p<0.003$). These results show that by the ages of four to five children attribute false belief to a robot but they do not attribute a mental verb to it. This means that, for children, the robot does not have thinking capabilities or thoughts in this situation.

In conclusion, we demonstrated that young children discriminate between a robot and a human in mentalizing when the mental verb was used in a question such as "think" in a False Belief Task. It seems to be difficult for young children to link the behavior of just searching and thinking in a robot.

Figures

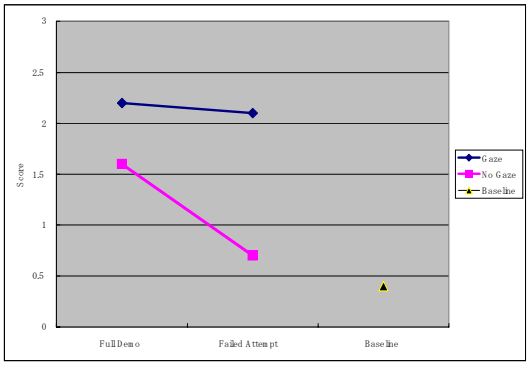

Figure 1: Results of Experiment 1.

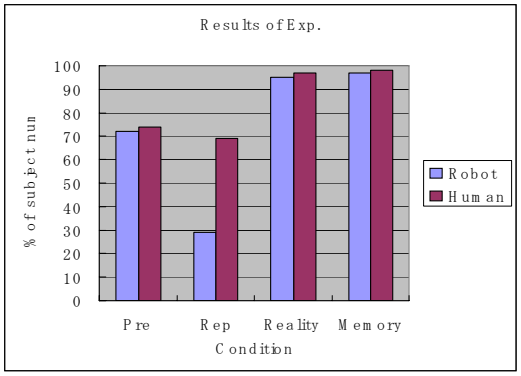

Figure 2: Results of Experiment 2.

Acknowledgments

These studies were supported by the ATR Intelligent Robotics and Communication Laboratory, and by a grant from JSPS (No: 13610087, 16500161) to Shoji Itakura.

References

Barr, R. & Hayne, H. (2000). Age-related changes in imitation: implications for memory development. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne, (Eds.), *Progress in infancy research*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development. 63* (4, Serial No. 255).

Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*, 447–458.

Frith, U., & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*, 459–473.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.

Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Science, 4*, 22–28.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*, 549–559.

Johnson, S. C., Booth, A., & O'Hearn, K. (2001). Inferring the goals of non-human agents. *Cognitive Development, 16*, 637–656.

Legerstee, M. (2001). Domain specificity and the epistemic triangle: The development of the concept of animacy in infancy. In F. Lacerda, C. von Hofsten, & M. Heimann (Eds.), *Emerging cognitive ability in early infancy*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Legerstee, M. (1991). The role of person and object in eliciting early imitation. *Journal of Experimental Child Psychology, 51*, 423–433.

Meltzoff, A. N. (1995). Understanding the intention of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838–850.

Phillipes, A. T. & Wellman, H. M. (in press). Infants' understanding of object-directed action. Cognition.

Poulin-Dubois, D. (1999). Infants' distinction between animate and inanimate objects: the origin of naïve psychology. In P. Rochat (Ed.), *Early social cognition.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Poulin-Dubois, D., Lepage, A., & Ferland, D. (1996). Infants' concept of animacy. *Cognitive Development, 11*, 1-16.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.

Exploring the Aesthetic Range for Humanoid Robots

David Hanson (david.hanson@utdallas.edu)
The University of Texas at Dallas PO Box 830688
Richardson TX 75083-0688

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. The author's preliminary experiments on human reaction to near-realistic androids appear to support this hypothesis.

Introduction

While Masahiro Mori's (1970) uncanny valley paradigm has dominated robotics design for over 30 years, there has been little formal challenge to the paradigm's premises. Is there really a strong, immutable relationship between the human realism and acceptability of robots? Is there an unavoidable discontinuity of acceptability (i.e., a valley) between abstract and highly realistic anthropomorphic depictions?

To answer this question, *human realism* needs to be clearly defined. 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 and dynamic appearance and contextual responsiveness (i.e., contingent interaction). Within each dimension, there are many sub-characteristics of realism, such as physical geometry, texture, and coloration, which will be constrained by human biology.

A humanoid figure may exhibit extreme realism in some characteristics while deviating from realism in others (e.g., a realistic face with a cartoon body). Alternately, the characteristics of a figure may evenly deviate from realism (e.g., a face and a body that are both slightly cartoonish). With so many ways to deviate from realism, and so many ways to modulate the 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 (Mori, 1970). In (Hanson et al., 2005), anecdotal examples indicated that there can be indeterminately many possibilities for aesthetic humanlike depictions that lie outside the curve of Mori's valley. This implies that human reactions to an anthropomorphic depiction are more strongly related to good or bad design than to its level of human realism.

This paper describes a series of preliminary tests that attempt to map out human reaction to robots that are nearly human-looking in appearance. The results of these tests appear to contravene the uncanny valley hypothesis. An alternative to the uncanny valley paradigm is then proposed.

Background

In recent years, neuroscientists and evolutionary psychologists have found abundant evidence that our tastes of beauty and ugliness are stamped into our nervous system (Rhodes and Zebrowitz, 2002), shaped 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 represent a primary obstacle course for social robot designers. Any "uncanny" perceptual phenomenon depends on these neural systems.

While studies indicate that we are much more sensitive to real human faces (Gauthier et al., 2000), the specific forms of beauty and ugliness inspire remarkably consistent human responses, regardless of their level of realism (Etcoff, 2000; Zaidel, 1997; Thomas and Johnston, 1995). The 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).

Universally, clear skin, well-groomed 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 universally considered 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).

It is well demonstrated that human aesthetic preferences transfer to nonhuman objects and beings (Norman, 1992; Kanwisher, 1997; Breazeal, 2002; Fong et al., 2003).

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 to be disturbing (Darwin and Ekman, 1872/1998; Etcoff, 2000). Facial forms akin to expressions of terror, psychosis, and subterfuge are also universally found to be alarming (Darwin and Ekman, 1872/1998; Ekman, 1970; Adolphs et al., 2001). These kinds of eerie signifiers are used in cartoons and art to depict villains or monsters. 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. Avoiding perceptual templates that trigger fear may help avoid the uncanny reaction, regardless of the level of realism.

But what about sensitivity to realism—is there any evidence that realism does make a difference? Studies do show that people are especially sensitive to the real human face (Tzourio-Mazoyer, 2002; Kanwisher, 1997; Kanwisher, 2000). We are much more sensitive to familiar faces and objects (Gauthier, 1998) and can more easily recognize such faces (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 change a real face from attractive to unattractive (Etcoff, 2000, p. 134). These sensitivities imply that more realistic faces trigger more demanding expectations for anthropomorphic depictions (Hanson et al., 2005).

Sending Robots in to Explore the Valley

Social robotics research has reapplied techniques of animatronics (entertainment robotics) in AI-driven robots, with notable examples including Cynthia Breazeal's collaboration with Stan Winston on the Leonardo robot, and Hiroshi Ishiguro's work with Kokoro Co., Ltd. on robots including the Repliee Q1. The author's robots continue this trend, being realistic in expression and interactive, but differ in that always some features put them in the region of the uncanny valley—for example, the back of the head is missing from the Philip K. Dick robot, exposing wires and mechanisms (see Fig. 1). These robots are intended to plumb the uncanny valley and challenge the premises of the paradigm.

Fig. 1. Philip K. Dick Android.

In addition to animated appearance, the author's robots engage with conversational speech, via AI-driven intelligent software using face tracking, face recognition, automatic speech recognition, and speech synthesis.

In November 2004, the author led two informal web-surveys that showed videos of two Hanson robots, animated with humanlike facial expressions. 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."

From June 2005 to November 2005, the Philip K Dick Android (with the back of the head missing) was shown in 3 public exhibitions, where people's behavior was observed and noted by curators. Following their interactions with the robot, people were given exit interviews. According to the observers, people who interacted with the robot appeared entertained, not disturbed or afraid. The robot held peoples' attention in conversation for many minutes and even hours. People held the android's hand while talking with it, and even spontaneously hugged the android at the end of the conversation. In the exit interviews, 71% said the robot was "not eerie," and 89% "enjoyed" interacting with the robot. These results seem to merit more formal experiments.

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 participants series of images of the Philip K. Dick android (PKD-A), Qrio, 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 the human models on which the robots were based. The human participants were asked to rank the images from 1 to 10, on several metrics: realism, appeal, eeriness, and familiarity.

Method

Participants. There were 25 participants, ranging in age from 18 to 77. The national origin of the participants was diverse, with 12 U.S. nationals, 8 south Asian, and 6 other. 12 Participants were male and 13 female. Participants were recruited in public thoroughfares on two college campus of the University of Texas at Dallas. All participants were volunteers and none received remuneration.

Procedure. In these experiments our control morph (see Fig. 2) was inspired by a morph used in the experimental work of Karl MacDorman, using the robots of Hiroshi Ishiguro, which contains a continuum of morphed images that elicit reactions from participants which follow the pattern predicted by the uncanny valley.

In our experimental morph, meanwhile, the morph images were designed with the intention of making them appealing and not eerie (see Fig. 3). If human participants reacted with consistently low-eeriness ratings, this would imply that the uncanny valley is avoidable, at least in the static domain.

Reaction to the control figures followed the pattern predicted by the uncanny valley theory (see Fig. 2). 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. 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 anthropomorphism across the range of realism, thus supporting the hypothesis of no inherent uncanny valley.

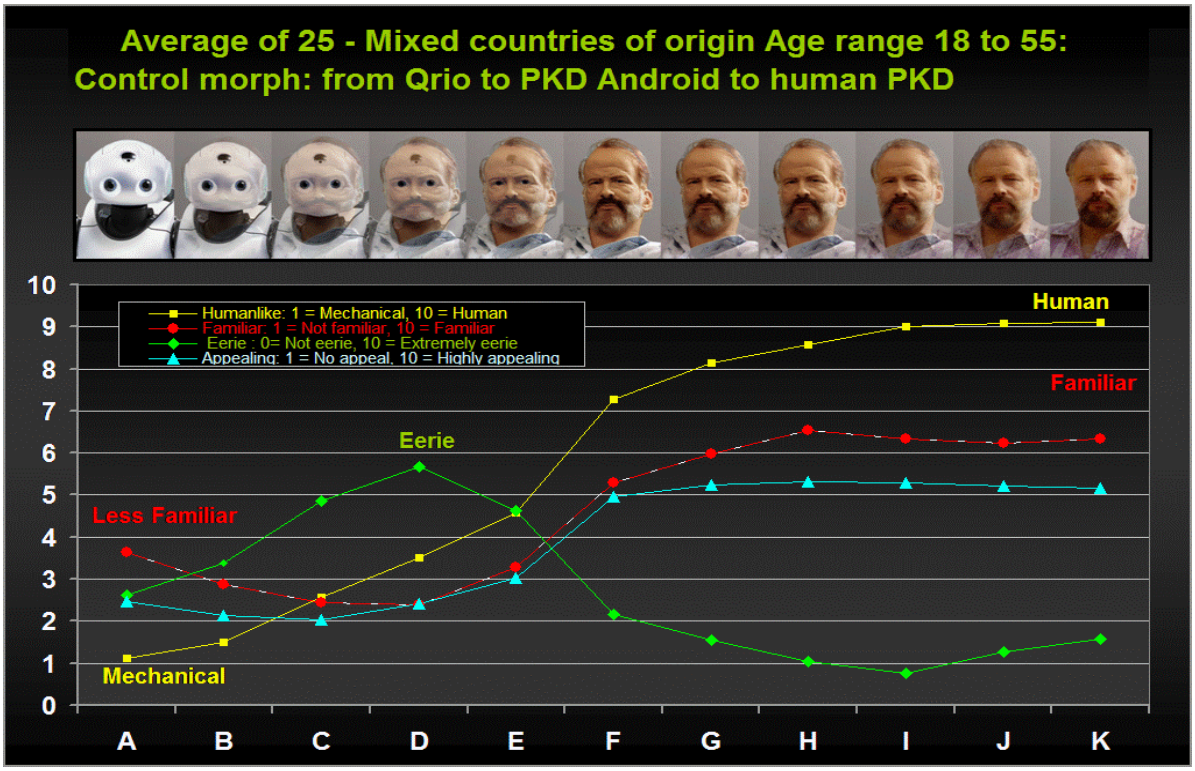

Fig. 2 (above). Uncanny morph: Qrio—android—PKD.

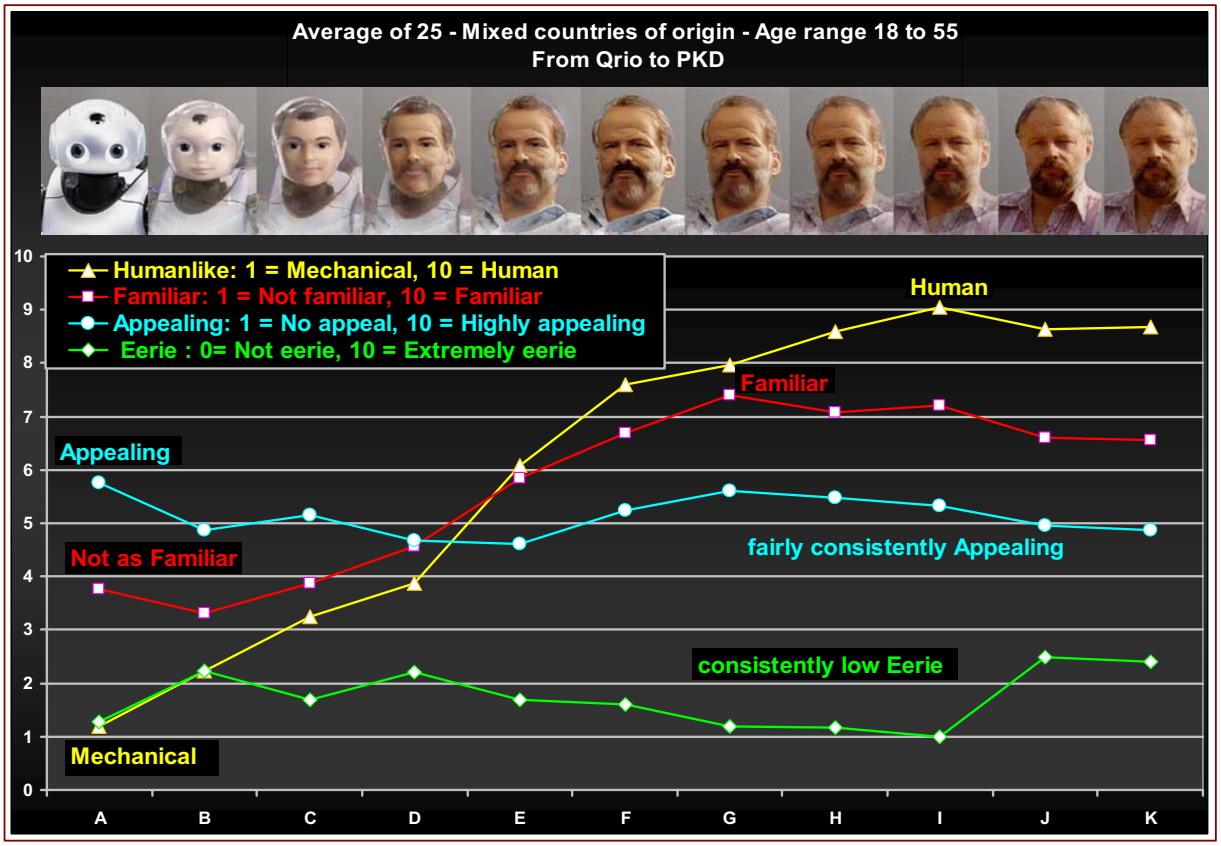

Fig. 3 (below). Not Uncanny morph: Qrio—android—PKD.

Thus the data results are not consistent with the uncanny valley hypothesis. Further experiments are merited. Future experiments should be animated, interactive, and with participants in the presence of actual robots. A range of realism should be produced in the robots.

Pursuing New Theory for Robot Design

We propose a preliminary a replacement paradigm for the Uncanny Valley. If the illusion of life can be created and maintained, the uncanny effects may be mitigated. It may be that any level of realism 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).

Conclusions 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.

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 in a humanlike appearance, can remind us of our own mortality. They also may imply dead matter impersonating humans, conveying the threat of an imposter. But, if we remove these flaws to make them friendly, attractive, and seemingly alive, then the level of realism may not matter.

Ultimately, good design can help to make robots lovable and part of 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, interaction, and cognition.

Acknowledgments

The author would like to acknowledge the help and guidance of Karl MacDorman, Alice Otoole, Thomas Linehan, and Dennis Kratz. The author especially thanks Amanda and Elaine Hanson for their support and assistance.

References

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

Breazeal C. (2002). *Designing Sociable Robot.*, Cambridge, Mass.: MIT Press.

Cunningham, M.R., Barbee A.P., & Philhower C. (2002). Dimensions of facial physical attractiveness: The intersection of biology and culture, in *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives*. Westport, Conn.: Ablex Publishing.

Darwin, C. & Ekman, P. (Ed.). (1998/1872). *The expression of the emotions in man and animals.* New York: Oxford University Press.

Dion K.K. (2002). Cultural perspectives on facial attractiveness, in *Facial attractiveness: evolutionary, cognitive, and social perspectives.* Ablex Publishing.

Eibl-Eibesfeldt, I. (1970). *Ethology: The biology of behavior.* New York: Holt, Rinehart and Winston.

Ekman, P. & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124-129.

Etcoff, N. (2000). *Survival of the prettiest*. New York: Anchor.

Fong, T., Nourbakhsh, I., Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42,* 143-166.

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

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

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

Hanson, D., Olney, A., Zielke, M., Pereira, A. (2005). Upending the uncanny valley, in *AAAI conference proceedings.*

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

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. *Cognitive Brain Research, 11,* 213-226.

Mori, Masahiro (1970). Bukimi no tani (the uncanny valley). *Energy, 7,* 33–35. (In Japanese).

Norman, D. (1992). *Turn signals are the facial expressions of automobiles*. Cambridge, Mass.: Perseus Publishing,..

Rhodes, G., & Zebrowitz, L.A. (2002). *Facial attractiveness: Evolutionary, cognitive, and social perspectives.* Westport, Conn.: Ablex Publishing.

Rubenstein A.J., Langlois J.H., Roggman L.A. (2002). What makes a face attractive and why: The role of averageness in defining facial beauty, in *Facial attractiveness: Evolutionary, cognitive, and social perspectives.* Westport, Conn.: Ablex Publishing.

Thomas, F. & Johnston, O. (1995). *The illusion of life: Disney animation* (Rev. ed.). Hyperion.

Tzourio-Mazoyer, N., De Schonen, S., Crivello, F., Reutter, B., Aujard, Y., Mazoyer, B. (2002). Neural correlates of woman face processing by 2-month-old infants. *NeuroImage, 15,* 454-461

Zaidel, D. W. (1997).Art and science. *Nature, 390,* 330.

An Inventory of Reported Characteristics for Home Computers, Robots, and Human Beings: Applications for Android Science and the Uncanny Valley

Christopher H. Ramey (cramey@flsouthern.edu)
Department of Psychology, Florida Southern College
111 Lake Hollingsworth Dr., Lakeland, FL 33801 USA

Abstract

The uncanny valley refers to a state of perceptual or cognitive experience at which an increasingly humanlike figure becomes strange, rather than increasingly more familiar or acceptable. This formulation, however, is predicated upon a clear notion of what *human likeness* is. Human likeness is a vague term that requires clarification if it is to be used as an independent variable in experimentation in android science. This paper inventories various reported characteristics of home computers, robots, and human beings. The purpose of this is to delimit empirical research in android science on those robot features necessary for the experience of the uncanny and for the formation of social relationships.

Uncannily Human

A world populated with humanlike androids could lead to "a new lifestyle with robots" (Ishiguro, 2005a, p. 5) for human beings. The question remains as to whether it is possible to design robots that are sufficiently humanlike to be assimilated into social relationships and the complex culture of human beings. An android is defined as "an artificial system that has humanlike behavior and appearance and is capable of sustaining natural relationships with people" (see Ishiguro, 2005a; MacDorman, Minato, Shimada, Itakura, Cowley, & Ishiguro, 2005). Indeed, one benchmark of the successful design of a social robot is this ability to sustain long-term, *natural* relationships with people (MacDorman & Cowley, in press; see also Ramey, 2005a, in press). One difficulty for the formation of social relationships, however, is the purported uncanny valley effect (Mori, 1970/2005).

The uncanny valley refers to a state of perceptual or cognitive experience at which an increasingly humanlike figure becomes strange, rather than increasingly more familiar or acceptable (see Mori's original [1970/2005] formulation; Mori, 2005).[1] Although Mori's principle has been around for over thirty years, little systematic empirical and theoretical research in the behavioral and engineering sciences has been conducted on the parameters of the uncanny valley effect (although see Chaminade, Hodgins, & Kawato, 2005; Ishiguro, 2005b; Keysers & Gazzola, 2005; MacDorman, 2005; Ramey, 2005b).

There are some researchers (e.g., Hanson, Olney, Pereira, and Zielke, 2005) who maintain that the uncanny valley does not exist or at least can be escaped through careful design; Hanson et al. (2005) recently concluded that participants rating real and non-real stimuli "showed no sign of the repulsion that defined the 'valley' of Mori's uncanny valley" (p. 7). They thus advocated that either the valley does not actually exist beyond Mori's intuition or is innocuous for robotics design and should no longer be avoided. To account for the difference in whether people experience the uncanny valley or not, other researchers (e.g., Ramey, 2005b) have argued that the uncanny valley should not be regarded as unique to the concerns of humanoid robotics. Rather, the uncanny valley effect is a member of a class of cognitive and perceptual states of uncertainty at category boundaries (i.e., *humans* and *robots*) for a novel stimulus (i.e., *humanlike android*). Freud (1919/2003), for example, advocated how personal knowledge and knowledge of context influence the experience of stimuli (e.g., as the uncanny or the mundane).[2]

At the very least, the notion of the uncanny valley and its relevance to android science is predicated upon a clear notion of what *human likeness* is (be it in the researcher's design and methodology or the participant's interpretation). 'Human likeness' is a vague term that requires clarification if it is to be used as an independent variable in experimentation. That is, to what extent something is humanlike will depend on what the stimulus is. (An extremely realistic humanlike foot on a robot is likely not as uncanny an experience as an extremely realistic humanlike face; of course, this intuition requires empirical analysis.) This paper is a preliminary inventory of various reported characteristics of home computers, robots, and human beings in order to delimit empirical research in android science to those features (their presence or lack thereof in design) necessary for the experience of the uncanny.

Method

Participants. Fifty-eight (N = 58) undergraduates (mean age 20.87 yrs) participated in this study for course credit. Participants were randomly assigned to one of four experimental conditions: Human-Computer (n = 13),

[1] Although as Freud (1919/2003) noted in his influential account of the uncanny, "beyond doubt… the word is not always used in a clearly defined sense" (p. 123).

[2] Freud (1919/2003): "Even when Pygmalion's beautiful statue comes to life, this is hardly felt to be uncanny… The false semblance of death and the raising of the dead have been represented to us as very uncanny themes. But again, such things are commonplace in fairy tales. Who would go so far as to call it uncanny when, for instance, Snow White opens her eye again? And the raising of the dead in miracle stories – those of the New Testament, for example – arouses feelings that have nothing to do with the uncanny" (p. 153).

Human-Robot (n = 15), Robot-Human (n = 13), and Computer-Human (n = 14).

Materials and Procedure. A target task incorporated in a brief demographic questionnaire was used. In this study, participants were asked to consider a typical instance of an item. As an example, they were to consider a typical "desk," not a specific "desk" that they could remember or were currently in. They then were to answer several questions.[3]

(1) What does it look like on the outside? Describe its appearance or visible parts.

(2) What does it do? What can it do? How does it behave?

Participants were then asked to provide ten characteristics of a typical instance of the mentioned item. The following example was provided:

(1) *Made of wood, flat, got 4 legs…*

(2) *Has me sitting in it, just sits there, pile stuff on it…*

After participants completed this task, they were asked to look over the characteristics in each column and consider them with respect to another item: "For example, pretend that I asked you to circle all features of a typical 'desk' that would also apply to a typical 'table.' For example, a 'table' can be made of wood, is normally flat, and has four legs. It also just sits there and can have stuff piled on it. It is also good because it helps you get good grades because you can study better with it." It is important to note that this second task item was not made known to participants before completing the first task.

After any questions were answered, participants in the Human-Computer, Human-Robot, Robot-Human, and Computer-Human conditions completed the target task in the manner of the orientation ask described above. As an example, consider the Human-Robot condition: Each participant first listed no more than ten characteristics of a *human being* and then circled those features that a *robot* also possessed.

Results and Discussion

Reported characteristics were transcribed and are summarized in the following Tables.[4] Table 1 displays the proportion of shared human, computer, and robot properties and attributions overall. Tables 2 and 3 display the frequencies of specific properties and attributions in the Human-Robot Condition and the Robot-Human Condition, respectively. Table 4 displays summary data with respect to face properties.

Two contrasts are immediately evident from Table 1. First, Robots have more in common with humans than computers do (Human-Robot vs. Human-Computer). Second, humans have more in common with robots than computers do (Robot-Human vs. Computer-Human). (Humans and computers do not appear to have much in common, and tables for Human-Computer and Computer-Human data are, thus, omitted from the present report.)

Table 1. Proportion of Shared Human, Computer, and Robot Properties and Attributions

Condition	Physical Appearance
Human-Computer (n = 13)	.06
Human-Robot (n = 15)	.42
Robot-Human (n = 13)	.31
Computer-Human (n = 14)	.11

Note. In the condition "Human-Computer," participants were asked to list features of humans and subsequently were asked to indicate which features were shared with computers. The n refers to number of participants. The labels Human-Robot, Robot-Human, and Computer-Human conditions follow on this logic.

There appears to be support for the idea that humanlike robots are not only possible but acceptable in terms of a mapping of physical appearance. There is a correspondence between a robot's appearance and a human being's physical appearance. That is, robots and humans seem to share major physical appearance features. However, there are two points of caution here. First, human beings have less in common with robots (Robot-Human) than robots have with human beings (Human-Robot). This makes sense given that robots (*a fortiori* androids) are presumably modeled after the human image. This may also point to a further asymmetry relevant to design. A human being may allow other entities to possess human physical features up to a point. However, if one notes the properties of that other entity, a human being will be less willing to identify with these *foreign* category properties. Human beings will remain steadfastly loyal to their own category's attributes. It is, thus, worth investigating what features are shared between robots and human beings (see Tables 2 and 3).

Table 2. Frequencies of Properties and Attributions of Human-Robot Condition

Condition	Physical Appearance	Shared
General comments		
2 arms, 2 legs, 1 torso, head	1	1
2 sexual types male or female	2	0
woman	1	0
male female parts	2	1
sexual dimorphism	2	0

[3] A third question was (3) What are its positive features? What are its positive contributions? These analyses are omitted from the present report because a similar question concerning negative features was not included originally.

[4] The Tables are necessarily influenced by the author's *a priori* notions of types and tokens, but they are presented *in detail* to provide researchers with a preliminary inventory of relevant reported characteristics. In addition, characteristics of the abilities and actions of home computers, robots, and human beings are omitted from the present report owing to space limitations.

height	1	1
tall or short	1	1
between 5 and 6 ft.	1	1
weight	1	1
stands on two legs	1	0
symmetric	1	1
belly button	1	0
body	2	2
head with eyes, nose, ears…	1	0
head	5	4
neck	1	0
shoulders	1	1
torso	1	1
front	1	0
back	2	1
beautiful or ugly	1	0
fat-skinny	1	1
can be different skin, race colors	2	0
curvy body nice shaped	1	0
breasts	1	0
nice butt	1	0
nice calf muscles	1	0
curvy	1	1
clothes	2	0
Total	*40 (.27)*	*18 (.29)*
Organic appearance		
hair	6	1
hair length	1	0
both male, female has hair	1	0
hair on top of head	2	0
long brown hair, straight	1	0
skin	4	0
Total	*15 (.10)*	*1 (.02)*
Mechanical appearance		
Total	*0 (0)*	*0 (0)*
Face		
face	3	2
2 eyes, a nose, and a mouth	1	1
eyes	3	3
2 eyes	6	2
2 eyes on the front of the head	1	1
blue/green eyes	1	0
ears	4	1
2 ears	3	0
2 ears on the side of head	2	1
mouth	7	2
smile	2	1
lips	1	1
lip gloss	1	0
nose	9	2
Total	*44 (.29)*	*17 (.27)*
Arms and legs		
arms	4	4
2 arms	9	8
legs	4	3
2 legs	8	6
muscular legs, defined	1	0
Total	*26 (.17)*	*21 (.33)*
Hands and feet		
2 hands and 2 feet	1	1
hands and feet	1	0
hands	2	2
2 hands	1	1
hand, arm	1	1
every hand has 5 fingers	1	0
fingers	2	0
nails	1	0
painted toe/finger nails	1	0
10 fingers and 10 toes	2	1
10 fingers	2	0
foot, leg	1	0
feet	2	1
2 feet	2	0
every foot has 5 toes	1	0
toes	2	0
10 toes	2	0
Total	*25 (.17)*	*6 (.10)*

Note. The *n* refers to the total number of properties minus uncodeable (Maximum *n* = 150; Total *n* = 150; Shared *n* = 63).

Table 3. Frequencies of Properties and Attributions of Robot-Human Condition

Condition	Physical Appearance	Shared
General comments		
"human like"	1	0
looks like human	1	1
masculine figure like a male	1	1
may be android-like	1	0
uniform	1	0
unnatural	1	0
may have bolts	1	0
a fan to cool itself	1	0
different sizes	1	1
a little shorter than me	1	0
big	1	1
chubby	1	1
compact	1	0
box-like	1	0
boxy	2	0
wide	1	0
solid	1	1
stiff	1	0
hard	1	0
smooth surface	1	0
stable	1	1
stands upright	1	1
square/round	1	0
breakable	1	1
geometric	1	0
green stripes	1	0
head	2	2

rotating head	1	0
slinky shaped neck to move	1	0
torso	1	1
no clothes	1	0
Total	*33 (.31)*	*12 (.35)*
Organic appearance		
Total	*0 (0)*	*0 (0)*
Mechanical appearance		
data board	1	0
disk-drive	1	0
places on structures to insert info	1	0
chips	1	0
electric	1	0
machine	1	0
mechanical	1	0
made of metal, metal parts, steel	10	0
grey/metallic	1	0
silver	1	0
shiny	6	2
possible plastic	1	0
wheels	4	0
nails	1	0
a metal piece	1	0
outlets for various plugs	1	0
screen	1	0
antennas, transmitter	2	0
lights	1	0
blinking lights	1	0
flashing lights	1	0
bright lights	1	0
has a keyboard	1	0
lots of buttons, tons	2	0
buttons, perhaps colors	2	0
Total	*45 (.42)*	*2 (.06)*
Face		
Scary face	1	0
Imitation of human face	1	1
Pair of eyes	1	0
Eyes, possibly	1	1
Lights as eyes, red	2	0
Optical apparatus	1	1
Rectangle mouth	1	0
Slits in side of head for ears	1	0
Total	*9 (.08)*	*3 (.09)*
Arms and legs		
arms and legs	1	1
arms	1	1
2 arms	2	2
arm-like structure	1	1
extending arms	1	1
2 hands, 2 legs	1	1
legs	2	2
2 legs	2	2
Total	*11 (.11)*	*11 (.32)*
Hands and feet		
10 fingers and toes	1	1
feet	1	1
2 feet	1	1
wheels, on	1	0
wheels for feet	2	0
hands	1	1
2 hands	1	1
tong-like hands to grab things	1	0
opposable thumbs	1	1
Total	*10 (.09)*	*6 (.18)*

Note. The *n* refers to the total number of properties minus uncodeable (Maximum *n* = 130; Total *n* = 108; Shared *n* = 34). There were 18 non-responses subtracted from the maximum *n*.

Given the inventory of properties and attributions (Tables 2 and 3), it becomes clear that certain *type* features (e.g., face features) parallel between human beings and robots, though their *tokens* are not equivalent (see Table 4).

Table 4. Frequency (and Proportion) of Face Properties

Condition	Face	Shared
Human-Computer (*n* = 122)	38 (.31)	0 (0)
Human-Robot (*n* = 150)	44 (.29)	17 (.27)
Robot-Human (*n* = 108)	9 (.08)	3 (.09)
Computer-Human (*n* = 120)	0 (0)	0 (0)

Note. The Face column refers to frequency (and proportion of Total *n*) of face properties. The Shared column refers to frequency (and proportion of all shared features) within that condition.

It is immediately clear that facial features are very important for the identification of human beings, whereas this class of properties is not so for robots or computers. Given that only 8% of robot features are facial features, whereas the comparable percentage for human beings is about 30%, *robots are not stereotypically defined by their face*. One might nonetheless expect that facial features attributed to robots would comprise a substantial amount of the features later attributed to human beings, but this is not the case. Robots are allowed to have human faces (Human-Robot, 27% of all shared features), but human beings are not allowed to have robot faces (Robot-Human, 9% of all shared features). Closer inspection reveals why this is the case. Robot facial features are quite different from human facial features. Participants' reported features like "scary face," "lights as eyes," and "slits in side of head for ears," and "rectangle mouth." The stereotypical robot face is a terrifying caricature of a human being's face.

General Discussion

The present paper investigated those features and attributions of human beings and robots that are stereotypically associated with them. This preliminary inventory is required because *human likeness* is a necessary variable for (a) the design of humanlike androids in android science and (b) the empirical and systematic investigation of variables relevant to the uncanny valley effect in robotics

research and android science. Future research should no longer rely on *intuitions* (cf. Mori, 1970/2005) but rather be based on empirical inquiry.

Acknowledgments

I am grateful to Elizabeth S. Lee for her assistance in data entry, as well as Evangelia G. Chrysikou for discussions concerning these data.

References

Chaminade, T., Hodgins, J., & Kawato, M. (2005). Exploring the uncanny valley: Behavioral and neuroimaging experiments. In *Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots*. Tsukuba, Japan.

Freud, S. (2003). *The uncanny*. (D. McLintock, Trans.; pp. 123-162). New York: Penguin (Originally published 1919)

Hanson, D., Olney, A., Pereira, I. A., & Zielke, M. (2005). Upending the uncanny valley. *Proceedings of the American Association for Artificial Intelligence (AAII) Conference*. Pittsburgh, PA.

Ishiguro, H. (2005a). Android science: Toward a new cross-interdisciplinary framework. In *Proceedings of CogSci-2005 Workshop: Toward Social Mechanisms of Android Science* (pp. 1-6). Stresa, Italy.

Ishiguro, H. (2005b). Lateral inhibition hypothesis for uncanny valley. In *Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots*. Tsukuba, Japan.

Keysers, C., & Gazzola, V. (2005). The neural basis of social cognitions and their responses to non-human agents. In *Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots*. Tsukuba, Japan.

MacDorman, K. F. (2005). *Memento Mori*: Are humanlike robots uncanny because they remind us of death? In *Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots*. Tsukuba, Japan.

MacDorman, K. F., & Cowley, S. J. (in press). Single white robot seeks *Homo sapiens* for long-term relationship: A new benchmark for robot personhood. In *Toward Psychological Benchmarks in Human-Robot Interaction*, a special session of *RO-MAN 06: The 15th IEEE International Symposium on Robot and Human Interactive Communication*: *Getting to Know Socially Intelligent Robots*. Hatfield, UK.

MacDorman, K. F., Minato, T., Shimada, M., Itakura, S., Cowley, S., & Ishiguro, H. (2005, July). *Assessing human likeness by eye contact in an android testbed*. Paper presented at the 20th Annual Meeting of the Cognitive Science Society. Stresa, Italy.

Mori, M. (2005). Bukimi no tani [The uncanny valley] (K. F. MacDorman & T. Minato, Trans.). Retrieved from http://www.theuncannyvalley.com (Originally published 1970; *Energy*, *7*(4), 33-35)

Mori, M. (2005). On the uncanny valley. *Humanoids-2005 workshop: Views of the uncanny valley*. December 5, 2005, Tsukuba, Japan.

Ramey, C. H. (2005a). 'For the sake of others': The 'personal' ethics of human-android interaction. In *Proceedings of CogSci-2005 Workshop: Toward Social Mechanisms of Android Science* (pp. 137-148). Stresa, Italy.

Ramey, C. H. (2005b). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In *Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots* (pp. 8-13). Tsukuba, Japan.

Ramey, C. H. (in press). Conscience as a design primitive in social robots. In *Toward Psychological Benchmarks in Human-Robot Interaction*, a special session of *RO-MAN 06: The 15th IEEE International Symposium on Robot and Human Interactive Communication*: *Getting to Know Socially Intelligent Robots*. Hatfield, UK.

Subjective Ratings of Robot Video Clips for Human Likeness, Familiarity, and Eeriness: An Exploration of the Uncanny Valley

Karl F. MacDorman
School of Informatics, Indiana University, USA

Abstract

Masahiro Mori observed that as robots come to look more humanlike, they seem more familiar, until a point is reached at which subtle deviations from human norms cause them to look creepy. He referred to this dip in familiarity and corresponding surge in strangeness as the *uncanny valley*. The eerie sensation associated with a mismatch between human expectations and a robot's behavior provides a useful source of feedback to improve the cognitive models implemented in the robot. Is the uncanny valley a necessary property of near-humanlike forms? This paper contributes to ongoing work in understanding the nature and causes of the uncanny valley by means of an experiment: 56 participants were asked to rate 13 robots and 1 human, shown in video clips, on a very mechanical (1) to very humanlike (9) scale, a very strange (1) to very familiar (9) scale, and a not eerie (0) to extremely eerie (10) scale. Contrary to earlier studies with morphs [MacDorman and Ishiguro, 2006], plots of average and median values for ratings on these scales do not reveal a single U-shaped valley as predicted by Mori's uncanny valley hypothesis [1970], although his hypothesis allows for some variation owing to movement. Robots rated similarly on the *mechanical versus humanlike* scale can be rated quite differently on the *strange versus familiar* or the *eeriness* scales. The results indicate that the perceived human likeness of a robot is not the only factor determining the perceived familiarity, strangeness, or eeriness of the robot. This suggests that other factors could be manipulated to vary the familiarity, strangeness, or eeriness of a robot independently of its human likeness.

Introduction

To build robots that at least superficially approach human likeness is leading to insights in human perception and face-to-face interaction. These *android* robots possess the physical presence that simulated characters lack, yet can be more perfectly controlled than any human actor, to isolate the factor under study. Even in experiments in which the android's responses are identical, we can observe how human responses vary according to their beliefs. For example, Japanese participants showed the same modesty with their eyes by averting gaze downward when interacting with an android as when interacting with a human interlocutor *if* they believed the android were under human control by telepresence [MacDorman et al., 2005].

In addition, androids provide an ideal testing ground for theories from the social and cognitive sciences because competing models can be implemented in an android and then tested by letting the android interact with people [MacDorman and Ishiguro, 2006]. Androids provide not only a platform for integrating techniques from science and engineering but also for studying the relationship between interaction and cognitive mechanism. Thus, they may one day provide an avenue for unifying the behavioral sciences and cognitive neuroscience.

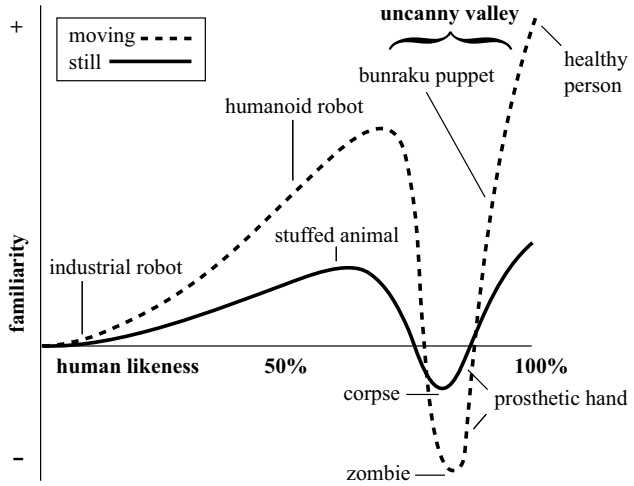

Figure 1: Mori hypothesized the relation between human likeness and perceived familiarity: familiarity increases with human likeness until an *uncanny valley* is reached caused by sensitivity to perceived imperfections in near-humanlike forms [Mori, 1970]. Movement, according to Mori, magnifies the uncanny valley.

Recent evidence indicates that androids are generally better able to elicit human-directed norms of interaction than less humanlike robots or animated characters [MacDorman and Ishiguro, 2006] [Cowley and MacDorman, 2006]. However, Mori [1970] observed a heightened sensitivity to defects in near-humanlike forms—an *uncanny valley* in what is otherwise a positive relationship between human likeness and familiarity (Fig. 1). Although Mori proposed that abstract human forms should serve as the principle for designing socially-acceptable robots, the uncanny valley can be seen more positively—for example, by indicating when a robot's responses do not rise to the expectations elicited by its human form. This provides useful feedback for improving the cognitive models implemented in the android (see Fig. 6).

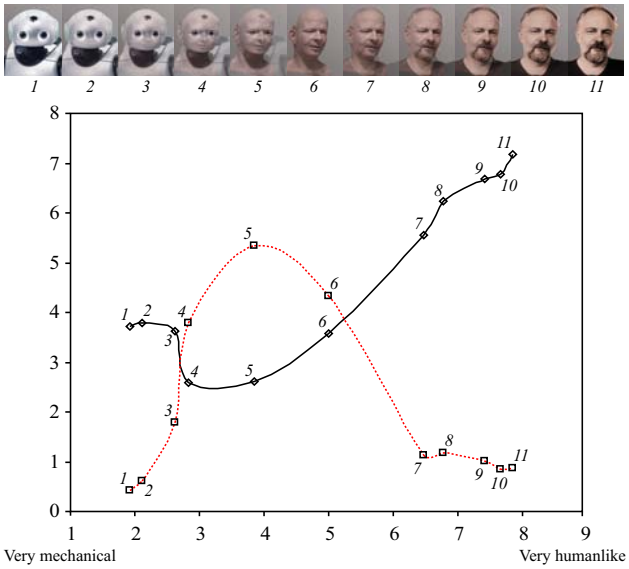

Figure 2: Average ratings for *strange versus familiar* (solid line) and *eeriness* (dashed line) were plotted against *mechanical versus humanlike* for images morphing from the robot Qrio the Philip K. Dick android to Philip K. Dick himself. The plots reproduce Mori's hypothesized uncanny valley and indicate a corresponding region of eeriness.

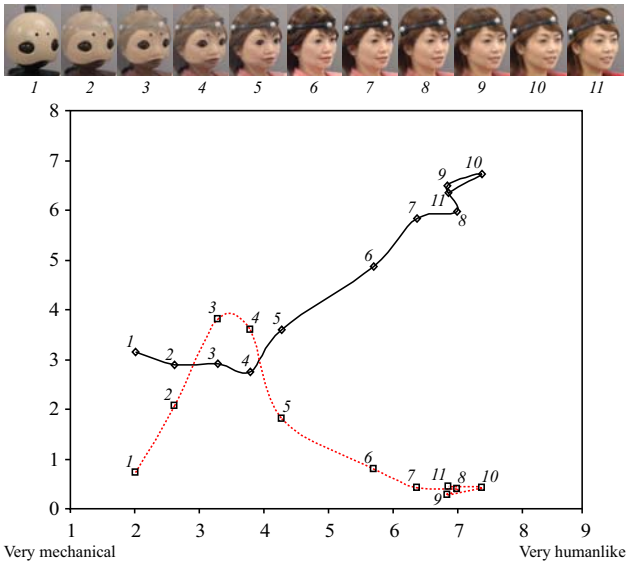

Figure 3: The same plots as Fig. 2 but for images morphing from the robot Eleviee to the android Repliee Q1Expo to the android's human model.

Plotting the uncanny valley

In a previous study, 45 Indonesian participants were asked to rate 31 images on a nine-point scale ranging from very mechanical (1) to very humanlike (9) and from very strange (1) to very familiar (9) [MacDorman and Ishiguro, 2006]. They were then asked to select eerie images and rate them for eeriness on a ten-point scale, ranging from slightly eerie (1) to extremely eerie (10).[1] Two sets of 11 morphed images were included among the 31 images.

Fig. 2 shows the plot of the average ratings on the *strange versus familiar* (solid line) and *eeriness* (dashed line) scale for the first set of images, which morphed from a photograph of the humanoid robot Qrio (left) to one of the Philip K. Dick android developed by Hanson Robotics (center) to one of Philip K. Dick himself (right). Fig. 3 shows the plot for a second set of images, which morphed from a photograph of the humanoid robot Eveliee (left) to one of the android Repliee Q1Expo (center) to Repliee's human model (right). The independent axis is the average rating on the *mechanical versus humanlike* scale. The plots reproduce Mori's posited uncanny valley (solid line) and indicate a region of eeriness in the same area (dashed line).

Experiment: Ratings of videos

The intention of the current study is to determine whether the uncanny valley is a necessary property of near-humanlike forms. Participants are presented with short video clips of a wide range of mainly android and humanoid robots engaged in various activities in different settings.

[1]Images that were not selected as eerie were rated 0.

Subjects. There were 56 Indonesian participants, 43 male and 13 female, of whom 13 were 17 to 20 years old, 36 were 21 to 25, 4 were 26 to 30, and 3 were 31 to 35. The participants were mainly university students, young professionals, and government workers. Participants were recruited from an Internet cafe and received two hours of free Internet access.

Procedure. Participants were asked on a computer-based questionnaire, in individual sessions, to rate 14 video clips, most of which were 30 to 60 seconds in length, on a nine-point *mechanical versus humanlike* scale, a nine-point *strange versus familiar* scale, and a ten-point *eeriness* scale. The scales ranged from very mechanical (1) to very humanlike (9), from very strange (1) to very familiar (9), and from not eerie (0) to extremely eerie (10). The video clips included a mobile robot (Pioneer II), a manipulator arm, seven humanoid robots (Rovovie-M3, HR-2, VisiON Nexta, Chronio, Robovie, Wakamaru, Asimo), two android heads (K-bot, Eva), two androids (Philip K. Dick, Repliee Q1Expo), and one human being. The video clips were presented in random order. For each video the three ratings were requested in random order. The direction of the scales was determined randomly for each question.

Results. Fig. 4 shows the plot of the average ratings on the *strange versus familiar* (solid line) and *eeriness* (dashed line) scale for a given average rating on the *mechanical versus humanlike* scale for the video clips of the 14 robots. There is no consistent valley shape when plotting familiarity against human likeness. Instead, there are oscillations in eeriness for robots that range from mechanical looking (1.96) to approaching very humanlike (8.57). The plot of *strange versus familiar* and *eeriness* are almost mirror images.

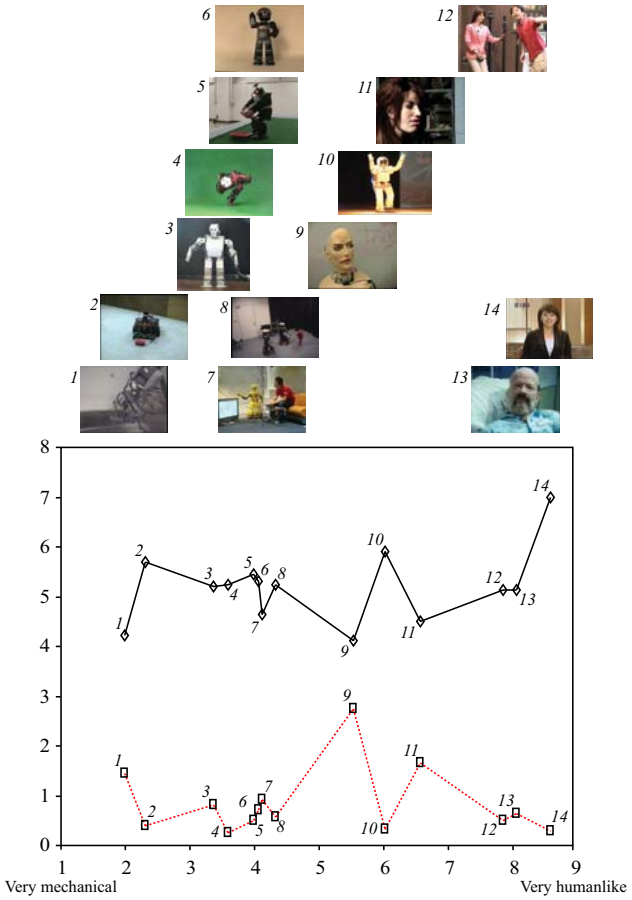

Figure 4: Based on the average ratings of 56 participants, 14 video clips are arranged from left, mechanical, to right, approaching very humanlike. The names of the robots are listed in Table 1. The solid line plots the relationship between perceived humanlikeness (on the *mechanical versus humanlike* scale) and perceived familiarity (on the *strange versus familiar* scale). The dashed line plots the relationship between perceived humanlikeness and eeriness. There is no single uncanny valley in the plot.

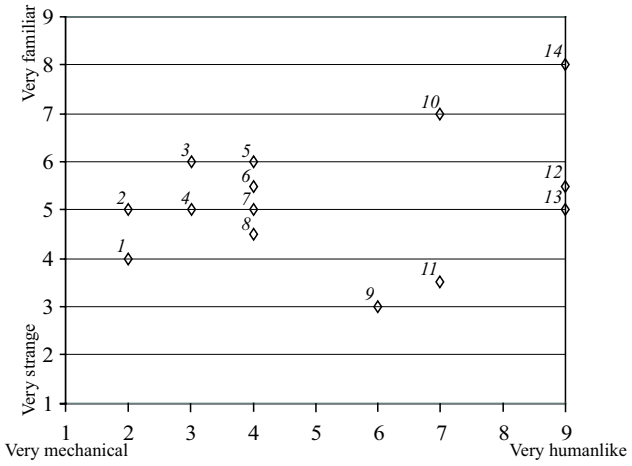

Figure 5: The median ratings on a *strange versus familiar* scale are plotted against the median ratings on a *mechanical versus humanlike* scale for the robots in Table 1 and Fig. 4. Robots with the same median value for human likeness can show quite different median values for familiarity.

Table 1: Median ratings of the 14 video clips on a very mechanical (1) to very humanlike (9) scale and a very strange (1) to very familiar (9) scale

No.	Name	mech. v hum.	strange v fam.
1	Manipulator	2	5
2	Pioneer II	2	4
3	HR-2	3	6
4	Robovie-M3	3	5
5	Nexta	4	6
6	Chronio	4	5.5
7	Wakamaru	4	5
8	Robovie	4	4.5
9	K-bot	6	3
10	Asimo	7	7
11	Eva	7	3.5
14	Human	9	8
12	Repliee	9	5.5
13	PKD android	9	5

The average standard deviation (SD) was 1.69 for the *mechanical versus humanlike* scale, 2.43 for the *strange versus familiar* scale, and 1.68 for the *eeriness* scale.

Median values were considered a more robust indicator of central tendency, especially given the high variance in the data and the the subjective nature of the questions. Table 1 lists the values for the *mechanical versus humanlike* and *strange versus familiar* scales. The median values for the *eerie* scale are not listed because they were all 0 except for K-bot whose median value was 1. Fig. 5 plots the median values for the *strange verus familiar* scale against the median values for the *mechanical versus humanlike* scale.

Table 1 and Fig. 5 show that video clips of robots that were rated as having the same median human likeness could have much more or much less median familiarity. For example, the median value was *very humanlike* (9) for the Philip K. Dick android, Repliee Q1Expo, and its human model, although the median value for the human model was *familiar* (8) but *neutral* (5) for the Philip K. Dick android and near neutral (5.5) for Repliee Q1Expo. The median values also represent Asimo as *somewhat familiar* and K-bot as *somewhat strange*, although Asimo was represented as more humanlike than K-bot. Thus, the depiction of Asimo in the video clip seemed to bump up its human likeness despite the fact that it does not have a humanlike face with skin, teeth, nostrils, pupils, and so on.

Discussion

The video clips exhibit a wide range of robots performing different actions in quite different contexts, sometimes with speech accompaniment. The results do not indicate a single uncanny valley for a particular range of human likeness. Rather, they suggest that human likeness is only one of perhaps many factors influencing the extent to which a robot is perceived as being strange, familiar, or eerie. This is an important result because it implies that factors other than human likeness could be

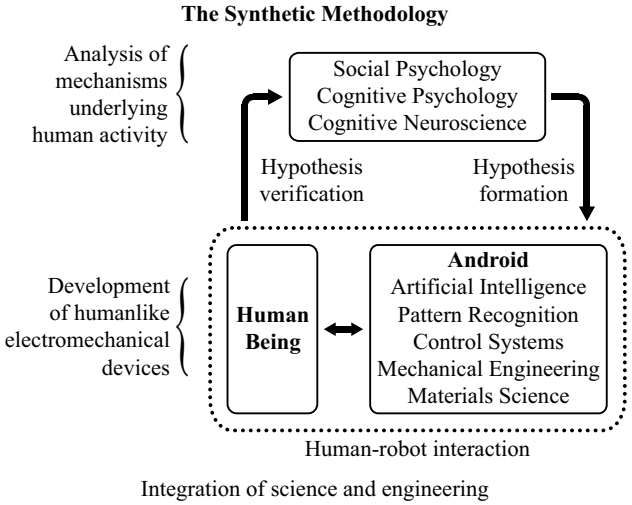

Figure 6: Ishiguro [2005] proposes a synthetic methodology for investigating human interaction that integrates science and engineering.

manipulated to overcome the uncanny valley. Of course, Mori [1970] [Mori, 1970] identified motion as one such factor, so it may be argued that his hypothesis could accommodate the results. But the variations in familiarity and eeriness for a given level of human likeness are not consistent with motion-induced magnifications of a valley shape.

So why does morphing from a mechanical-looking robot to an android and then to its human model produce an uncanny valley in still images, as shown in Fig. 2 and 3? One possible explanation is that the morphs provide a more gradual and consistent change with less extraneous variation. For example, as the images were still, variations in movement and speech were not an issue. We might expect to find uncanny valleys in more controlled experiments that vary appearance or movement along fewer parameters.

The notion that the uncanny valley can be escaped by varying factors unrelated to human likeness is consistent with an experiment performed by Hanson [2006] using morphs. Although he found that morphing from a mechanical-looking robot to an android produced a valley in a familiarity scale and in an appealing scale and a peak in an eeriness scale, these effects were greatly reduced by tuning the morphs. Thus, without making a morph more or less humanlike, Hanson was largely able to design around the uncanny valley. His technique was to adjust the appearance of the uncanny morphs toward the cuter features of a doll.

Conclusion

The results of the experiment suggest that human likeness is only one factor determining the familiarity, strangeness, and eeriness of a robot. This offers the hope that other factors modulating these qualities will be uncovered. MacDorman and Ishiguro [2006] have documented a number of possible explanations for the uncanny valley, ranging from expectation violation and cognitive paradoxes [Ramey, 2005] to evolutionary aesthetics [Etcoff, 1999] and pathogen avoidance. As the validity of these explanations comes under scientific scrutiny, design principles will appear that engineers can use to develop robots with desirable aesthetics. Whether it is indeed desirable to build a robot that is appealing or unnerving will depend on its purpose and the setting in which it is used. Given the success of the horror genre, it is clear that eeriness is not always considered a bad thing.

Acknowledgments

Much appreciation goes to Heryati Madiapuri for recruiting participants and conducting the experiments in September, 2005, to Yuli Suliswidiawati, for reviewing the experimental procedures, and to Z. A. Dwi Pramono and Christopher H. Ramey for serving on the ethics advisory panel and providing comments on the experimental design. Thanks also go to David Hanson, Almir Heralik, Takayuki Kanda, Hiroshi Ishiguro, Takahashi Tomotaka, ActivMedia Robots, ATR, Honda, Kokoro Co., Ltd., Mitsubishi, and Vstone for the robots.

References

[Cowley and MacDorman, 2006] Cowley, S. and MacDorman, K. (2006). What baboons, babies, and tetris players tell us about interaction: A biosocial view of norm-based social learning. *Connection Science*.

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

[Hanson, 2006] Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. In *Proceedings of the ICCS/CogSci-2006 Symposium: Toward Social Mechanisms of Android Science*, Vancouver, Canada.

[Ishiguro, 2005] Ishiguro, H. (2005). Android science: Toward a new cross-disciplinary framework. In *CogSci-2005 Workshop: Toward Social Mechanisms of Android Science*, pages 1–6, Stresa, Italy.

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

[MacDorman et al., 2005] 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.

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

[Ramey, 2005] Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In *IEEE-RAS International Conference on Humanoid Robots*, Tsukuba, Japan.