Assessing Human Likeness by Eye Contact in an Android Testbed

Karl F. MacDorman,[1,2] **Takashi Minato,**[1] **Michihiro Shimada,**[1]
Shoji Itakura,[3] **Stephen Cowley,**[4] **and Hiroshi Ishiguro**[1,5]
[1]Department of Adaptive Machine Systems and [2]Frontier Research Center
Osaka University, Suita, Osaka 565-0871 Japan
[3]Department of Psychology, Kyoto University
Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501 Japan
[4]School of Psychology, University of Hertfordshire,
College Lane, Hertfordshire AL10 9AB United Kingdom
[5]Intelligent Robotics and Communication Laboratories, ATR
2-2-2 Hikaridai, Keihanna Science City, Kyoto, 619-0288, Japan

Abstract

The development of robots that closely resemble human beings enables us to investigate many phenomena related to human interactions that could not otherwise be investigated with mechanical-looking robots. This is because more humanlike devices are in a better position to elicit the kinds of responses people direct at each other. In particular, we cannot ignore the role of appearance in giving us a subjective impression of social presence or intelligence. However, this impression is influenced by behavior and the complex relationship between it and appearance. As Masahiro Mori observed, a humanlike appearance does not necessarily give a positive impression. We propose a hypothesis as to how appearance and behavior are related and map out a plan for android research to investigate this hypothesis. We then examine a study that evaluates the behavior of androids according to the patterns of gaze fixations they elicit in human subjects. Studies such as these, which integrate the development of androids with the investigation of human behavior, constitute a new research area fusing engineering and science.

Introduction

Progress is underway to develop humanoid robots that can support rich, multimodal interaction [Kanda et al., 2004], and we may expect to see adequate competencies within the next decade for brief exchanges in stereotyped situations. However, these robots will be of substantially less value if because of their appearance, ordinary people are unable to accept them as a social presence. Studies of person-to-person interaction in psychology and other fields generally take our human form for granted. This leaves us to assume that our everyday impressions of sociality are a subjective phenomenon arising from our *interactions* with other people.

However, the importance of a humanlike appearance has yet to be discounted, and there are a number of reasons why it might be significant. We have a range of biomechanical structures that have evolved or been adapted to express volition, intention, and emotion: Our eyes indicate the direction of gaze, which supports joint attention and other interactive responses; our faces and vocal tract are populated by scores of muscles involved in controlling facial expressions and the voice; and our bodies are animated by gestures and other meaningful acts. In addition, we are highly sensitized to these biomechanical structures and have developed specialized brain centers to interpret them, including those implicated in identifying faces [Farah et al., 2000], detecting faces [Kanwisher et al., 1997] and hands [Downing et al., 2001], and recognizing emotion.

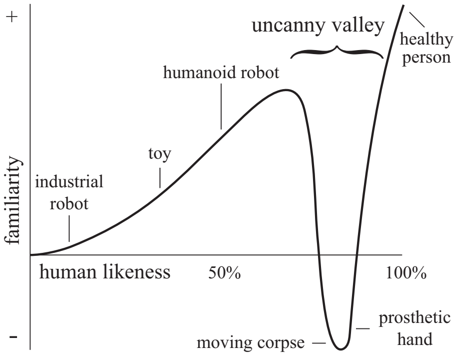


Figure 1: Mori's *uncanny valley* for animated objects [Mori, 1970].

Honed by evolution and experience, our most highly developed model of a social other is our model of other people. If we cannot accept humanoid robots as a social presence—even socially "competent" ones—because they do not look human, this is something robotics engineers need to know and plan for accordingly. This need has strongly motivated robotics engineers to learn something about us as people and how the human form—and deviations from it—affect our perceptions and reactions. Simply put, what makes something a social presence? Is it mainly its behavior, or is there instead some complex interplay between appearance and behavior?

Running counter to the view that we should build robots that look like people—what we call *androids*—is Masahiro Mori's hunch that our goal should instead be stylishly designed robots, because robots that look too human might be disturbing [Mori, 1970]. Mori proposed that our sense of familiarity increases as robots appear more human until an *uncanny valley* is reached at which subtle defects in human likeness appear repulsive (Fig. 1). The impression would not be unlike that of a moving corpse.

Only recently is Mori's hunch materializing into a research program for understanding the uncanny valley [Minato et al., 2004]. The effect of similarity can be separated into the effects of appearance and behavior, since both interdependently influence human-robot interaction. Goetz et al. [Goetz et al., 2003] observed a synergistic effect in our evaluation of an interaction when appearance and behavior are well-matched. Figure 2 averages graphs derived from Mori's uncanny valley hypothesis [Mori, 1970] and the hypothesis that appearance and behavior should be well-matched.

While one may argue that a robot's degree of human likeness should be adjusted to ensure that people neither place too few expectations on it (perhaps treating it like a piece of furniture) nor too many, it is unlikely that either appearance or behavior can be reduced to a single dimension. The temporal dimension is also missing from the figure. People tend to habituate to even an uncanny appearance, and behavior results in the development of relationships over time. In addition, a person in a spacesuit may not look so different from the sort of humanoid robots that we might expect to lie near the first peak (e.g., Honda's Asimo or Sony's Qrio); however, we would expect people to evaluate the movement of the astronaut more positively. Nevertheless, Minato et al. [Minato et al., 2004] may be right to hypothesize that a robot's uncanniness can be mitigated by its behavior, if the behavior closely resembles that of a person.

However, the uncanny valley can also be seen in a positive light. It indicates we are unconsciously applying our model of other human beings to the android—a model more demanding than the one we apply to mechanical-looking humanoid robots. It is an artifact of a mismatch between the more stringent demands of the human model that has been elicited and, in our progress toward human likeness, some vestigial sensory or sensorimotor data that does not match it.

Mori gave the example of shaking a prosthetic hand on a dark night [Mori, 1970]: We may feel uneasy if, after seeing it, we expect a human hand but then discover from its movement, feel, and temperature that it is instead a mechanical prosthesis. In this example, there is a mismatch both in terms of time and modality: The largely nonconscious expectations that at first fit the visual data cannot be reconciled with the tactile data. If, owing to the humanlike appearance of an android, human subjects are applying their model of other people, it becomes easier in an experimental setting to determine when the robot's behavior conforms to or deviates from the norms that people apply to each other.

This kind of knowledge is clearly of value to robot engineers in generating more natural movements. But it is also relevant to researchers in the cognitive and social sciences because it concerns human behavior. The implication is that an android may be able to go beyond the limits of mechanical-looking robots to serve as a testbed for theories about human behavior and for understanding the relationship between control mechanisms and social interaction.

However, the shape of the uncanny valley cannot be explained merely in terms of the elicitation of expectations about other people and the violation of those expectations because, as we near 100% human likeness, more human-directed expectations would come "on line," increasing our sense of familiarity, while fewer of them would be violated. Thus, we may posit at least two different models: one that rewards anthropomorphism in general with feelings of familiarity, and a second model that punishes deviations from human norms in figures that seem very human.

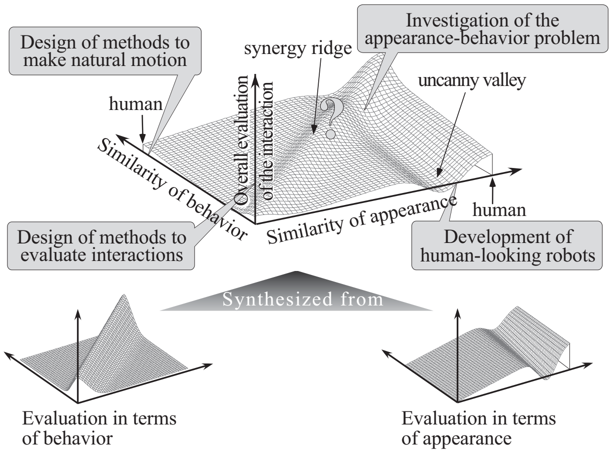


Figure 2: The extended uncanny valley and a map for investigating it.

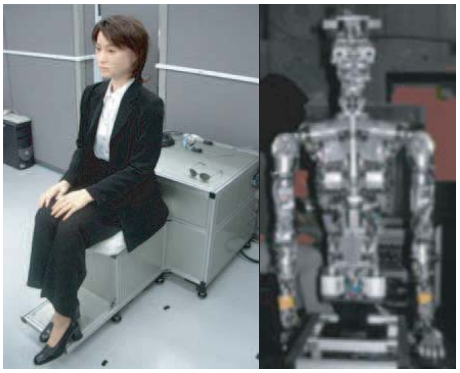


Figure 3: The android *Repliee Q1.*

Android Research Map

It may seem the final goal of android development should be to realize a device whose appearance and behavior cannot be distinguished from those of a human being (in other words, a device that could pass the Total Turing Test at T3 [Harnad, 1989]). However, since there will always be subcognitive tests that could be used to detect subtle differences between the internal architecture of a human being and an android [French, 1990] [French, 2000], an alternative goal would be to realize a device that is nearly indistinguishable from human beings in everyday situations. In the process of pursuing this goal, our research aims to investigate principles underlying interpersonal cognition and communication. Three main research issues define the axes of Fig. 2.

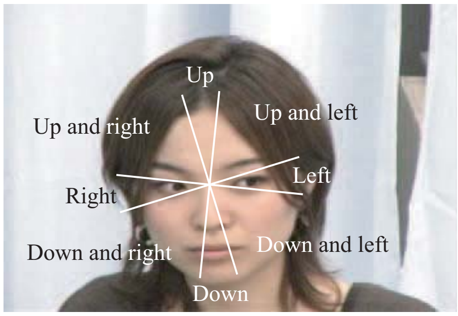
A method to evaluate human-robot interaction. Human-robot interaction can be evaluated by its human likeness. Therefore, it is necessary to compare human-human and human-robot interactions. Qualitative measures include the semantic differential (SD) method. Quantitative measures include statistical descriptions of a person's largely nonconscious behavior including gaze behavior, interpersonal distance, and vocal pitch. These observable responses reflect cognitive processes we might not be able to infer from answers to a questionnaire. We are studying how a human subject's responses reflect the humanlike quality of an interaction and how they relate to the subject's mental state.

Implementing natural motion in androids. To understand the kinds of motion that give a natural impression, the android precisely mimics a person's movement. We then monitor how a human subject's interaction with the android degrades as we remove some aspect of the android's motion. A straightforward way to animate the android is to design a sequence of control commands. However, this is difficult because the android has many degrees of freedom. Another method is to copy the motion of a human model as measured by a motion capture system. Most methods that use a motion capture system assume a human body has the same kinematics as a robot in calculating the robot's joint angles [Nakaoka et al., 2003]. However, because human and robot kinematics differ, there is no guarantee the robot's motion as generated from the angles will resemble human motion. Therefore, we need a method to ensure that the motions we see at the surface of the robot resemble those of a human being.

In particular, a human motion may be decomposed into dominant motions and fine motions that are contingent on the dominant motions. While a dominant motion may often be consciously initiated, it will result in fine motions that are largely nonconscious. For example, when raising a hand, a person's shoulder and waist may also move to keep balance. Breathing may become more rapid during physical exertion. These motions are considered important if an android is to closely resemble a person. We are studying methods to decompose human motion into dominant, contingent, and autonomic motions in addition to methods to map human motions to the android by means of an appropriate decomposition.

The development of humanlike robots. We have developed several androids we are currently using for experimentation. The android used in the experiments described in this paper is *Repliee Q1*, shown in Fig. 3, which was developed to realize humanlike motion. Repliee Q1 has 31 degrees of freedom in the upper body. The android can generate various kinds of micromotions such as the shoulder movements typically caused by human breathing. Silicone skin covers the head, neck, hands, and forearms. The compliance of the air actuators makes for a safer interaction. Highly sensitive tactile sensors mounted just under the android's skin enable contact interaction.

(a) The experimental setup



(b) The eight eye directions as coded

Figure 4: In the experimental setup, a human or android questioner interrogated Japanese college-aged students (a). Eight averted gaze directions were coded as shown in (b) as was eye contact.

A study of appearance and behavior

Breaking eye contact during thinking

In the evaluation of a human-robot interaction, methods of evaluating a human subject's (largely nonconscious) responses provide a complementary source of information to insights gleaned from a questionnaire or focus group. This paper examines the subjects' gaze behavior. Gaze behavior in human-robot interaction can be compared to the gaze behavior in human-human interaction, which has been studied in psychology and cognitive science.

In terms of gaze behavior, people generally make eye contact by looking with their right eye at the interlocutor's right eye. While thinking, people often break eye contact (avert their eyes from the interlocutor.) Three main theories explain this behavior:

- Arousal reduction theory

  This theory claims that people break eye contact while thinking to reduce arousal and to focus on the problem [Gale et al., 1978] by eliminating distractions.

- The differential cortical activation hypothesis

  This hypothesis states that brain activation induced by thinking tasks leads individuals to shift their gaze away from the central visual field [Previc and Murphy, 1997].

- Social signal theory

  This theory claims that people break eye contact to inform others that they are thinking.

If breaking eye contact were a social signal, we would expect it to be influenced by the interlocutor. Psychological researchers have reported experimental evidence to support the social signal theory [McCarthy et al., 2001, McCarthy and Muir, 2003]. We report an experiment that compares subjects' breaking of eye contact with a human versus android interlocutor.

Experiments

Human-human conversation It has been reported that Canadian subjects break eye contact longer for questions that require thinking with a preference for the upper-right direction; however, there was no directional bias for questions that do not require thinking [McCarthy et al., 2001, McCarthy and Muir, 2003]. The preference for the upper-right direction is considered to be the effect of a social signal. Although differential cortical activation is considered to cause a downward-averting gaze, people look up and to the right during interaction with others to avoid looking downward, which is considered to be negative behavior in Canada.

Subjects. The subjects in all experiments were Japanese college students within the 18–25 age range.

Procedure. Subjects sit opposite a human questioner (Fig. 4(a)). The questioner was a female Japanese college student. The subjects' eye movements are measured while they are considering the answers to questions posed by the questioner. There are two types of questions: *know questions* and *think questions*. Subjects already know the answer to know questions (e.g., "How old are you?") but not to think questions as these questions force the subject to derive the answer (e.g., "Please tell me a word that consists of eight letters."). The subjects were asked 10 know questions and 10 think questions in random order. Their faces were videotaped and the gaze direction was coded beginning from the end of the question to the beginning of the answer. Figure 4(b) shows the coding scheme for the eight averted directions, the ninth direction being eye contact.

Results. As in the Canadian study, the Japanese subjects tended to make less eye contact for think questions (27% of the time on average, SD=19%) than know questions (40%, SD=14%). Student's t-Test (two tails, two-sample unequal variance) is 0.1354. However, in contrast to the Canadian study, Japanese tended to avert their eyes downward for both know and think questions. Table 1 lists the duration of gaze in the eight directions shown in Fig. 4(b) (also see Figure 5(a)). From the figure, the duration of averting eyes is longer for think questions; however, there is almost no directional bias. Therefore, unless the signal is not present in Japanese culture, the social signal theory is not supported by the comparison between the know and think questions.

Human-android conversation We hypothesized that if the way in which eye contact is broken while thinking acts as a social signal, subjects will produce different eye movements if the interlocutor is not humanlike or if the subjects do not consider the interlocutor to be a responsive agent. Conversely, if eye movement does not change, this supports the contention that subjects are treating the android as if it were a person, or at least a social agent.

Procedure. We then conducted an experiment with eight subjects that was almost identical to the one described in the previous section except we substituted Repliee Q1 for the human questioner and told subjects the android was controlling its own behavior. Repliee Q1 resembles a young woman (Fig. 3). A speaker embedded in the android's chest produced a prerecorded voice. Micromotions such as eye and shoulder movements were implemented in the android to make it seem more humanlike.

At first the experimenter sitting beside the android explained the experiment to the subject to habituate the subject to the android. The android behaved as an autonomous agent during the explanation (e.g., it continuously made slight movements of the eyes, head, and shoulders while occasionally yawning). It seemed the subject believed the android to be asking questions autonomously, although questions were manually triggered by an experimenter seated behind a partition.

Results. The subjects tended to make much less eye contact for think questions (25%, SD=21%) than know questions (57%, SD=28%). Student's t-Test is 0.02326, which is very significant. Table 2 and Fig. 5(b) show the results.

If we assume the downward directional preference in human-human interaction is a social signal, the weaker

Table 1: Human Questioner

Gaze direction	think	know
Eye Contact	27%	40%
Up	3%	3%
Up and left	5%	6%
Left	16%	13%
Down and left	9%	2%
Down	20%	20%
Down and right	5%	5%
Right	11%	9%
Up and right	3%	2%

Table 2: Android: "Autonomous" or "Operated"

Gaze direction	think	know	think	know
Eye Contact	25%	57%	16%	54%
Up	4%	2%	4%	0%
Up and left	3%	2%	5%	3%
Left	16%	7%	10%	7%
Down and left	7%	8%	18%	7%
Down	14%	8%	27%	15%
Down and right	14%	7%	10%	5%
Right	7%	5%	5%	5%
Up and right	9%	4%	4%	3%

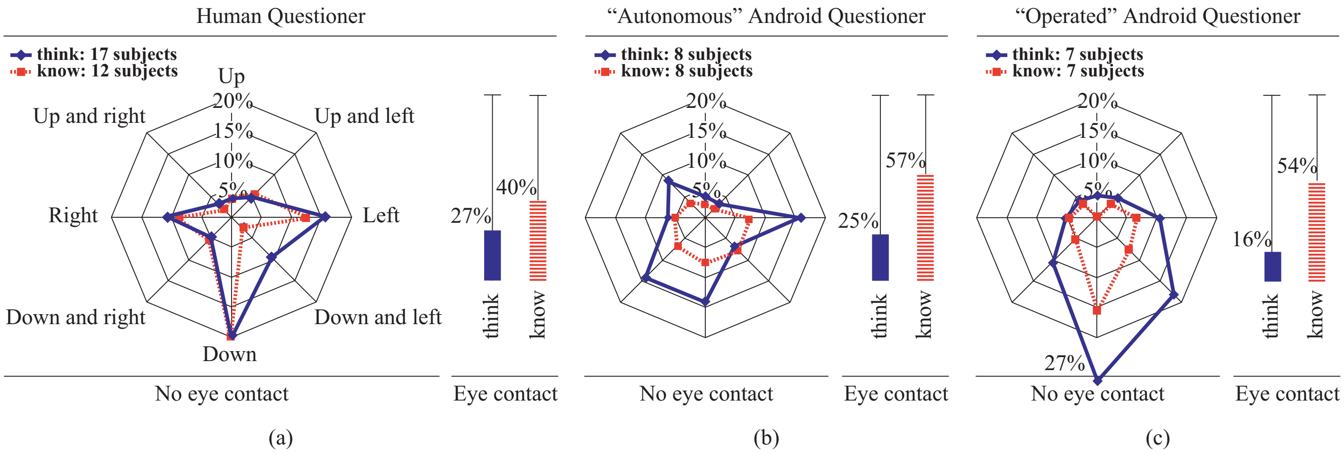


Figure 5: (a) For the 17 subjects in the "think" question experiment with the human interlocutor, the average amount of time spent making eye contact was 27%, and the average amount of time spent averting the eyes downward was 20%. For the 12 subjects in the "know" question experiment, the average amount of time spent making eye contact was 40% and the average amount of time averting the eyes downward was 20%. (b) The think question and know question experiments were repeated with an android with the subjects being told the android was "autonomous" and (c) with the subjects being shown how the android was operated by an experimenter.

preference in human-android interactions may suggest the subjects were not treating the android as an interactive agent. To check this reasoning, we conducted another experiment with seven subjects in which the subjects were told that the android is *not* autonomous and that an experimenter triggers the android to ask questions. We predicted the downward directional preference would decrease because the subject no longer considers the android to be a social agent.

Results. The subjects tended to make much less eye contact for think questions (16%, SD=15%) than know questions (54%, SD=21%). Student's t-Test is 0.002406, which is highly significant. Table 2 and Figure 5(c) shows the results.

Contrary to our expectation, the downward preference increased. This may be because the subjects were sending a social signal to the experimenter. However the downward preference is much less pronounced for the android believed to be autonomous. The difference in the gaze bias with respect to the different questioners suggests that breaking eye contact depends on the interlocutor. It also suggests that the sociality of the autonomous android is lower than the human questioner for the subjects. Conversely, breaking eye contact can be an evaluation of the android's appearance and behavior. We must, therefore, investigate which aspects of appearance and behavior influence human gaze behavior.

Under the condition that the subject believes the android to be human-operated, we consider the subjects interacted with the experimenter through the android and the relation between the experimenter and subject is different from that between the human questioner and subject. The difference may indicate that breaking eye contact has meaning as a social signal. However, the sample size is too small and the variance in response too great to make a detailed interpretation.

Discussion

Gaze may have a function not unlike how Cowley describes the interpersonal role of prosody in conversations [Cowley, 1994]. In other words, it may operate as a (predominantly subconscious) social response resulting from the experience of living in a culture. On this view, gaze is constrained not only neurally and socially but is itself primitive behavior that falls under the dual control of two brains. As a speaker acts to change her cognitive state, the interlocutor's gaze automatically serves as a cognitive resource. In this sense, gaze is intrinsic to *epistemic action* [Kirsh and Maglio, 1994]. When we talk, we affect each other's gaze just as we affect real-time patterns in each other's speech. Thus, gaze is not so much a "signal" or outer reaction to an environmental stimulus (e.g., a think question) as a way of contextualizing by drawing on experience in ways that are likely to be beneficial to the gazer.

On this view, Canadians and Japanese behave differently[1] because they have come to orient to different norms or, in population terms, have adopted different gaze practices. This joint activity is not standardized to anything like the extent that would be needed for it to be meaningfully described by a grammar. Except in such extreme cases as staring, gaze is not normative in the sense that we can formalize its function, say, in terms of social codes. Gaze is, however, norm-based. This is because deviations from common practice will take on meaning in relation to both circumstances and an individual's current perspective. Thus, any social group will be sensitive to distinctions between marked (preferred) and unmarked (disfavored) gaze responses.

---

[1] To put it more precisely, their behavior may be characterized by different probability distributions.

To test the merit of this line of speculation, it is necessary to consider not only the focus of gazes but also their duration and time sequence: The main question is whether eye movements function as signals that provide information about whether the subject is, for example, thinking or whether they function as prompts, probes, and teasers whose timing and other qualities are themselves the information shared among parties in closely-coordinated interaction. They may in fact function as both at once.

If gaze functions epistemically, a person (or android) can use invariants that develop over a life history to model *what action affords*. If Canadian the person knows (or acts as if she knows) that looking often (but not continuously) at the right eye is likely to give a sense of whether the interaction is proceeding normally and that looking up and to the right will not invite social sanction. In enacting this behavior, one person can use another as a cognitive resource that emits heterogeneous bundles of cues (e.g., as exhibited by the timing of mutual gaze). These cues prompt real-time adjustments as both individuals engage each other interpersonally while orienting what they do around norm-based behavior that has stabilized for some time period (but will typically disappear). Some of these will be pragmatic actions; others will be future epistemic actions [Kirsh and Maglio, 1994]. This can lead, among other things, to the development of largely nonconscious forms of activity that are likely to achieve affective reward. What may be crucial though is that epistemic action can use variable bundles of external features whose value depends on culturally-stabilized patterns.

Conclusion

This paper proposed a hypothesis on how appearance and behavior are related and mapped out a plan for android research to investigate it. The study on breaking eye contact during thinking was considered from the standpoint of the appearance-behavior problem. In the study, we used the android to investigate the social signal theory and obtained evidence differing from previous psychological experiments in human studies. Furthermore, it was found that the breaking of eye contact can be an evaluation of an android's human likeness.

References

[Cowley, 1994] Cowley, S. (1994). The role of rhythm in conversations: A behavioural perspective. *Language & Communication*, 14:353–376.

[Downing et al., 2001] Downing, P., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293:2470–2473.

[Farah et al., 2000] Farah, M., Rabinowitz, C., Quinn, G., and Liu, G. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology*, 17:117–123.

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

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

[Gale et al., 1978] Gale, A., Kingsley, E., Brookes, S., and Smith, D. (1978). Cortical arousal and social intimacy in the human female under different conditions of eye contact. *Behavioral Processes*, 3:271–275.

[Goetz et al., 2003] Goetz, J., Kiesler, S., and Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The Twelveth IEEE International Workshop on Robot and Human Interactive Communication*, Lisbon, Portugal.

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

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

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

[Kirsh and Maglio, 1994] Kirsh, D. and Maglio, P. (1994). On distinguish epistemic from pragmatic action. *Cognitive Science*, 18:513–549.

[McCarthy et al., 2001] McCarthy, A., Lee, K., and Muir, D. (2001). Eye gaze displays that index knowing, thinking and guessing (poster). In *The Annual Conference of the American Psychological Society*.

[McCarthy and Muir, 2003] McCarthy, A. and Muir, D. (2003). Eye movements as social signals during thinking: Age differences. In *Biennial Meeting of the Society for Research in Child Development*.

[Minato et al., 2004] Minato, T., Shimada, M., Ishiguro, H., and Itakura, S. (2004). Development of an android robot for studying human-robot interaction. In Orchard, R., Yang, C., and Ali, M., editors, *Innovations in Applied Artificial Intelligence*, pages 424–434, Berlin.

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

[Nakaoka et al., 2003] Nakaoka, S., Nakazawa, A., Yokoi, K., Hirukawa, H., and Ikeuchi, K. (2003). Generating whole body motions for a biped humanoid robot from captured human dances. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 3, pages 3905–3910.

[Previc and Murphy, 1997] Previc, F. H. and Murphy, S. J. (1997). Vertical eye movements during mental tasks: A reexamination and hypothesis. *Perceptual and Motor Skills*, 84(3):835–847.