Long-term relationships as a benchmark for robot personhood

Karl F. MacDorman, *Member, IEEE,* and Stephen J. Cowley

***Abstract*— The human body constructs itself into a person by becoming attuned to the affective consequences of its actions in social relationships. Norms develop that ground perception and action, providing standards for appraising conduct. The body finds itself motivated to *enact* itself as a character in the drama of life, carving from its beliefs, intentions, and experiences a unique identity and perspective. If a biological body can construct itself into a person by exploiting social mechanisms, could an electromechanical body, a robot, do the same? To qualify for personhood, a robot body must be able to construct its own identity, to assume different roles, and to discriminate in forming friendships. Though all these conditions could be considered benchmarks of personhood, the most compelling benchmark, for which the above mentioned are prerequisites, is the ability to sustain long-term relationships. Long-term relationships demand that a robot continually re-create itself as it scripts its own future. This benchmark may be contrasted with those of previous research, which tend to define personhood in terms that are trivial, subjective, or based on assumptions about moral universals. Although personhood should not in principle be limited to one species, the most humanlike of robots will be best equipped for reciprocal relationships with human beings.**

I. The problem space

It is problematic to define one of the goals of humanoid robotics as the creation of artificial human beings, because *human being* is, at least partly, a *biological* category and *robot* is an *electromechanical* category. This paradox can be sidestepped if we redefine the goal as the creation of an artificial person, while defining *person* with language that is free of speciesism (i.e., the presumption of human superiority) [1][2]. If it is not *Homo sapiens* DNA that makes us persons, what is it? Unless you are an extreme idealist, other people are assumed to have consciousness. So one criterion for judging an entity's personhood is whether it is conscious.For example, if an alien species arrived on the earth—one of obvious intelligence—in determining whether to afford it rights, ethicists would focus on whether it were conscious and had feelings, thoughts, hopes, and fears, not whether it were human.

Nevertheless, to lose consciousness, as in a deep sleep, is not to lose personhood. Also, we consider many species of animals to be conscious but not necessarily to be persons. And still, given the problem of "other minds," we may have doubts about human simulacra, such as android robots [3], ever being conscious, regardless of how humanlike their behavior might be. Since there is no consensus on why human beings are conscious, at this stage it is hard to predict whether machines could be conscious. However, if a robot looks and appears to act human, it may be hard to resist treating it as a fully conscious person (see Fig. 1) [4].

Fig. 1. Without understanding anything, the android robot Repliee Q1 (center) could give an impression of human presence by mimicking human autonomic and attentional movements. Repliee Q1 was a joint effort of Hiroshi Ishiguro's Intelligent Robotics Laboratory at Osaka University and Kokoro Co., Ltd.

Brief operational tests of intelligence, such as the Turing test [5], in which a computer is expected to pretend to be human, are both too easy and too difficult. They are too easy, because a mindless program can fool ordinary people into thinking it is human [6][7][8]. On the other hand, they are too difficult, because a clever judge can devise questions that no computer however brilliant could answer as a human being would—namely, questions designed to tease apart its subcognitive architecture [9]. Clearly, the Turing test, whether conducted in its original form across a teleprinter [5] or in its more recent robotic incarnations [10][11], suffers from speciesism [1][2]. This may be one reason for the test's waning significance [12]. Tests that limit the length of interaction or demand human verisimilitude fail to define the concept of *person* operationally.

So what is a person? Daniel Dennett, as summarized by Ross and Dumouchel [13] provides a useful starting point from the standpoint of a third-person analysis of interaction:

> Biological systems of the *H. sapiens* variety turn themselves into people—socially embedded teleological

K. F. MacDorman is Associate Professor in the Human-Computer Interaction program of the School of Informatics at Indiana University.

S. J. Cowley is Senior Lecture in the Department of Psychology at the University of Hertfordshire.

> selves with narrated biographies in terms of these very beliefs and desires—by taking the intentional stance toward themselves. They can do this thanks to the existence, out in the environment, of public languages that anchor their interpretations to relatively consistent and socially enforced rules of continuity.... [T]hey are incentivized to narrate themselves as coherent and relatively predictable characters, and to care deeply about the dramatic trajectories of these characters they become... [People] are partly constituted out of their social environments, both in the networks of expectations that give identity to them as people, and in the fact that the meanings of their own thoughts are substantially controlled by semantic systems that are collective rather than individual. They are thus not identical to their nervous systems, which are indeed constituted internally. (pp. 264-265)

According to Cowley [14][15], infant brains develop under the dual control of infant and caregiver toward rewarding patterns of interaction. The infant is led to a practical understanding of intentional activity through the caregiver's over-interpretation of the infant's actions, which are partially guided by affective reward from the caregiver.[1] As the infant comes to integrate the physical and affective consequences of actions, the infant unwittingly adapts to—and even changes—norms that can be described on many levels, ranging from the biomechanical to the intersubjective and the cultural [16]. Infants become self-implicating as they develop skill in assessing the likely affective consequences of their actions, including their vocalizations in context. By the age of one, they become self-regulating, integrating body movements and syllabic patterns to influence the caregiver's actions. The infant and caregiver's affective appraisals during interaction imbue circumstances with meaning. In a given circumstance, the infant's conformance to and deviation from norms will take on meaning for both the caregiver and the infant from each of their perspectives [17]. These perspectives are partly defined by their affective appraisals. Affect, thus, provides grounding for activity and intentional co-action, which includes symbols, gestures, and other forms of real-time human expression.

How could robots construct themselves as people? Granted, we might be able to build robots that are already like people, but that does not simplify the problem. The importance of critical cycles to the learning of language and the social significance of actions notwithstanding, the processes that bring an adult person into being may not be so different from those that take that person into the future. You are not the same person you were 10 or 20 years ago. Personal identity follows a course of continual, though gradual, recreation, as our beliefs, circumstances, and relationships change with time. Real people develop mutual expectations in a relationship that are specific to the individuals involved. Furthermore, having experiences that match a real past is important to maintaining the trust and authenticity underpinning successful relationships. To build a robot that lacks the ability to develop its identity and beliefs—or at least *simulated* beliefs—in tandem with evolving social relationships is to develop a robot that is stuck in a moment in time.

[1]We would argue that the adult does not necessarily need a mental intention to engage in intentional activity, although when queried the adult may be able to give a first person report of the intention as a kind of rationalization after the fact.

II. Benchmarks for social interaction

The tendency in human-robot interaction literature is to consider only behavioral benchmarks that focus on the experience or behavior of the human beings with whom the robots are interacting. One problem with this approach is that analysis focuses on the human participants or on the interaction, which is to say, the human being and robot taken as a system. But there is no such thing as a generic human being that can be used in a standardized benchmark. There are only individuals with particular likes, dislikes, relationships, and idiosyncratic ways of thinking, feeling, and behaving.

This is not to deny the existence of norms. They do exist, but many levels of description may be applied to them, including the individual, inter-individual, group, inter-group, and socio-cultural [18]. Treating human beings as if they were generic emphasizes biological commonalities and ultimately devalues individual differences in human beings and robots alike. If a robot is ever to be its *own* person with its *own* point of view, a different approach must be employed. Norms can be violated or not violated, but when they are violated or not violated in a given relationship, group, or culture, it has meaning for the individuals involved. Behavior means nothing unless analyzed in its present and historical context. A human being can "dance like a robot," which could be amusing or embarrassing depending on the situation, but it would not make the human being any less of a person or any more of a robot. Likewise, for a robot to dance like a human being does not make it any more of a person, unless its responsiveness to the context indicates otherwise.

It is important to understand that *contingency* is inherent in human relationships. People develop intertwined biographies that exhibit mutual expectations and ways of interacting that are unique. A particular woman, say, Jane Smith, is more than just a human being; she is the resident heart surgeon at St. John's hospital, the devoted wife of Daniel Smith, and so on. The way she speaks to her daughter Kate will be different from the way she speaks to anyone else. Robots too need to be able to grow into different roles and relationships. Robots and simulated characters are already able to imitate human behavior and to process human languages at a rudimentary level; nevertheless, it seems easier to form a relationship with a language-less dog than with these machines. What does the dog have that the machines lack? The dog has a highly rich set of sensors and actuators, feelings and responsiveness to contingent behavior and human affective displays, an ability to behave in context, and—perhaps most importantly—

the ability to maintain *individual* relationships.[2] It is not surprising to find that a pet dog has particular ways of responding to every member of its human family. Moreover, these ways of responding develop and change over time. Dogs are responsive to operant conditioning and much else. And clearly a person's capacity to relate to other persons far exceeds that of a dog. Therefore, *an ability to develop and maintain individual relationships may be a useful benchmark for human-robot interaction*, if the goal is to make robots that are more like persons. The robot must be able to "go on" inside the relationship [19][20], not simply be able to interact in some brief exchange. Often the most important thing is not to follow any specific rules (e.g., of etiquette) but to simply figure out how to "fit in" with people [21][22] while maintaining one's own identity, values, and aspirations, which can also involve resistance, argument, and negotiation.

In sum, just as a human body manages to turn itself into a person, so must a robot body. To do this, presumably, the robot body will avail itself of some of the same kinds of social mechanisms that human bodies do. The robot should be able to develop or at least maintain its own identity, beliefs, and perspective on life. It should be able to maintain relationships with people that are unique, individual, and personal. According to the context, it should also be able to assume different roles, if doing so is in its own interest. No doubt attuning to norms on many levels will be essential to this. The robot may even form friendships, but not arbitrarily. It must pick its own friends, just as we do, and, of course, it may have enemies too.

III. Benchmarks proposed by other researchers

A number of psychological benchmarks of human-robot interaction are to be found in the literature (see Table I). Although most of the benchmarks could be useful for appraising personhood depending on how they were applied, none of them fundamentally confronts what it means to be a person.

The benchmarks suffer from three main kinds of problems: First, some of the benchmarks are simplistic. Either the benchmark is a simple method of quantifying data instead of a true benchmark or, although the benchmark may have some merit, its operational definition does not capture its full significance. For example, a quantitative analysis of *interaction properties* [23] is not a benchmark *per se*; however, this technique can be very useful in analyzing interaction data, thus providing a method of defining other benchmarks operationally.

The second kind of problem is that some of the benchmarks presume that universal moral values exist and that the quality of being human can be measured by one's degree of conformance to those values. Unfortunately, these benchmarks do not concern what makes an entity a *person*, but rather what makes a person a *better* person from a particular moral standpoint. In other words, they do not measure *how much* of a person an entity is, but rather *how good* a person that entity is. Even more worrisome is the fact that the benchmark proposer is treating the criteria as morally absolute, not culturally relative, when it is the proposer's cultural worldview that sets the criteria for goodness.

TABLE I

A classification of psychological benchmarks. Trivial benchmarks have simplistic operational definitions. Moral benchmarks are based on a presumption of universal moral values. Subjective benchmarks are highly sensitive to the participant's viewpoint or relationship to the robot.

Benchmark	Trivial	Moral	Subjective	Reference
autonomy (neglect tolerance)	✓			[24][25]
autonomy (social)		✓	✓	[26][27]
cooperation		✓	✓	[28]
engagement	✓		✓	[29][30]
human awareness	✓			[31]
imitation	✓			[27]
interaction properties	✓			[23]
intrinsic moral value		✓	✓	[27]
moral accountability		✓	✓	[27]
persuasiveness	✓		✓	[32]
privacy		✓	✓	[27]
politeness			✓	[33]
reciprocity			✓	[27]
self-awareness	✓			[31]
trust			✓	[34]

The third kind of problem affecting most of the benchmarks is that they are highly subjective. Some of these benchmarks are very sensitive to the study participant's viewpoint or ideology. For example, to some people it is "common sense" that machines are just machines and thus unworthy of moral consideration. To these people, it would not matter if the machine behaved like a toaster or like a human being.[3] Other subjective benchmarks are very sensitive to the role of the robot in the participant's life and the particular nature of their relationship. For example, privacy concerns with a sexual surrogate robot would be different from those with a co-worker robot.

A. Trivial operational definitions

The degree of *engagement* between a robot and the people around it could be a useful psychological benchmark for appraising the robot's personhood, depending on how engagement is defined operationally and the period of time over which it is measured. If engagement were maintained over a long period of time, and without any exogenous reason, such as a need to work together, it would be a sign of a relationship. However, it would be a mistake to define engagement solely in terms of how quickly a robot can capture a person's attention and how long it can hold it [29][30], because these metrics do not reflect the mutual contingency that is essential to closely coordinated social interaction. Indeed, these metrics can be applied just as easily to passive media. Nielson, for example, makes a business

[2]Of course, in relationships with dogs, people may fall into a certain amount of projection or over-interpretation of their behavior.

[3]The director of the flesh fair in the film *AI: Artificial Intelligence* expressed this attitude.

of measuring which television programs and commercials people are paying attention to, but these programs and commercials do not engage their audience in interaction. By these metrics, a robot that simply replayed a stand-up comedy routine might prove more engaging than a robot capable of interpersonal response.

Additional metrics mentioned by Steinfeld et al. [31] include *autonomy, human awareness,* and *self-awareness.* Autonomy, however, should not be defined narrowly as *neglect tolerance* [24][25]. A self-repairing robot on an uninhabited planet could be considered highly autonomous in the sense of neglect tolerance, though it might be completely unsociable. Kahn [26] argues that autonomy should be understood as "highly social, developed through reciprocal interactions on a microgenetic level, and evidenced structurally in incorporating and coordinating considerations of self, others, and society" [27]. Likewise, human awareness should not be limited to the ability to detect the presence of human beings in the surroundings. And finally, it is simplistic to define self-awareness as the ability to understand one's own limitations, monitor one's self, and detect and recover from faults [31]. One's awareness of self needs to be at least partially defined in terms of one's awareness of others.

Kahn et al. [27] also proffer *imitation* as a psychological benchmark, but not in the sense of blind imitation. Although a child may imitate behaviors blindly during an initial projective phase, following Baldwin [35], Kahn et al. argue that these actions take on meaning in a subjective phase, and finally during an ejective phase the child expresses acquired subjective understandings to others. Although imitation, or mimesis, is a possible benchmark, it may be only one of many aspects of co-action and the mechanisms involving the affective grounding of communicative behavior in relationships. The factors that motivate imitation are of primary interest. Persons act for their *own* reasons and not merely out of a natural tendency to imitate. In mother-child interactions, imitation is only part of what is going on. Cowley [36] found that Zulu mothers could quiet their three-month-old babies with a hand signal reinforced by a smile after the baby quiets. Probably what the infant is 'imitating' is not the mother but his or her own behavior on some prior occasion, which was rewarded by the mother's voice and smile. The process exploits the child's sensitivity to contingencies and the mother's sensitivity to the child's sensitivity, prompting children to develop self-directed behavior.

B. The presumption of universal moral values

A number of proposed 'psychological' benchmarks which supposedly indicate what a human being is turn out to be moral benchmarks indicating what a good person is, such as Kahn et al.'s concept of autonomy [27]: "Now, autonomy means in part independence from others. For it is only through being an independent thinker and actor that a person can refrain from being unduly influenced by others (e.g., by Neo-Nazis, youth gangs, political movements, and advertising)." But if autonomy is applied as a benchmark in this way, Neo-Nazis and other gang members might be deemed less than human, when in fact they are fully human but intent on living up to moral standards that differ from our own. By our standards, they may be wicked; but by their standards, we too may be wicked. The ethical thread running through these benchmarks suggests that what makes people human is their humanity. But human beings are not only capable of greater humanity than other animals but also greater atrocities, at least when judged from the standpoint of a given moral perspective.

Other psychological benchmarks have similar problems, such as whether people would find humanoid robots to have *moral accountability* for their actions or would grant them *intrinsic moral value* [27]. Would people feel that it is okay to isolate a robot, to treat it as a slave, to buy and sell it, or to cause it physical harm? Or would they feel that the robot should be afforded human rights and dignity? The problem is that this question is culturally, historically, and personally relative. Thinking on human rights is a moving target. Even the Old Testament permitted slavery, a practice which still continues today in some countries. It may be useful to compare human perceptions of robots to perceptions of other creatures that are considered less than human. People who work in a slaughterhouse may feel no different about killing lambs than they do about stapling boxes (some may even enjoy it), while some animal rights activists may feel animals have rights equal to those of human beings [1][2]. If robots find themselves in the same psychological terrain as animals, a realm somewhere between persons and objects, differences in worldviews will need to be carefully controlled for in experiments. Arguably, there are some people who only value the lives of people with whom they have close personal attachments. If human participants in a study with robots are of the opinion that "people matter" and they are willing to decide whether a robot is a person based solely on its outward behavior and appearance, intrinsic moral value becomes a useful benchmark. But if they believe people are expendable, the results will become skewed.

Although we may appeal to moral 'universals,' psychologically speaking, the universal may not be these particular moral values, but simply the fact that each person has a worldview against which to measure the conduct of self and others. Worldviews are somewhat arbitrary and strongly affected by upbringing, culture, and peer attitudes. In a street gang, or a tribe of cannibals, it could be considered 'cool' to be what we might call depraved. Different cultures and religions have different 'virtues' and 'vices,' exhibiting different worldviews. When our conduct measures up to our worldview, we enjoy self-esteem [37][38]. When it does not, we may become paralyzed with fear, guilt, or shame. In this sense, our worldview is a reflection of our identity and how we gauge our own self-worth. But the view that human beings have intrinsic moral value may not be a psychological universal, since what is valued is to some extent culturally relative. While most people would certainly like to believe that there are universal moral values, just as they would like to believe that goodness will prevail and evil will be punished, the world does not necessarily works that way. It

may be just a story we tell ourselves to escape from nihilism: the idea that life has no intrinsic purpose, value, or meaning; that there is no God or universal truth supporting our moral values; and that death only brings personal annihilation.[4] And so we create myths that imbue life with meaning and offer a literal or symbolic transcendence of mortality, either through our continued existence in the afterlife or through identifications with things that do go on: our children, our country, our religion, our shared principles, our basketball team, our professional output, and so on.

While civil libertarians may tend to shun bigots, or religious zealots, those who break the edicts of their religion, the universal that Solomon et al. [38] found is that whatever our worldview is, when we are subconsciously reminded of our own mortality (e.g., through priming), we are more likely to favor those who uphold our worldview than those who criticize it. And this defense is lessened by self-esteem. The defense mechanism appears to be based on our organism's biologically programmed desire for self-preservation, wherein self is not simply defined by our physical existence but constructed from all that we identify with.[5] This mechanism seems to be an example of a human psychological universal, and not any particular worldview, such as one that favors intrinsic moral value. So the human psychological universal may not be the particular moral values human beings share but the fact that we, as a society, construct and, as individuals, possess moral values that are influenced, in characteristic ways, by self-esteem, identity, and the desire for self-preservation.

C. Highly subjective benchmarks

All moral benchmarks of human robot interaction are subjective insofar as the evaluative standard is derived from a cultural worldview that is, to some extent, arbitrary. Anthropologists, psychologists, and human ethologists may try to isolate universal values, including values derived from biological universals; however, there has been little attempt to do so in the literature on human-robot interaction. It is worth considering other psychological benchmarks that are not so overtly moral in nature though still highly subjective.

Kahn et al. [27] propose *privacy* as a possible benchmark for human-robot interaction. The mother in the film *AI: Artificial Intelligence* felt disturbed when her android son walked in on her in the bathroom. The sense of encroachment she felt indicates that she was experiencing the android to be, in some sense, human. Presumably, she would not have had the same feeling if, say, a robot vacuum cleaner had entered the bathroom. While respect for a robot's privacy and disapproval of its encroachments seem like good indicators that a person views a robot as a social entity, there are many other equally valid metrics, such as *politeness* [33]. Respect for privacy may be just one particular nexus of human norms and, as such, depend on the values of particular individuals, families, and cultures. Privacy, for example, was less important to children who lived together on a kibbutz than in ordinary Jewish families, but the lack of privacy of the kibbutz did not appear to be a source of neuroses [40]. On the contrary, the kibbutz children tended to have fewer neuroses.

In Japan there is a wide variety of humanlike sex dolls, which no doubt will be mechanized in the near future. Having a sexual surrogate robot walk in on one in the shower may have an entirely different significance with relation to feelings about privacy as compared to other kinds of robots. Feelings surrounding a robot's encroachment on one's privacy have more to do with one's *relationship* to the device than whether it like a person. The way the relationship has developed, habituation, the appearance of the robot, and the social matrix in which it is embedded, all would mediate feelings related to privacy.

Other highly subjective benchmarks include *persuasiveness* [32] and *trust* [34]. But it is not clear that a lack of these qualities makes a person any less of a person, though it may make a person less convincing or less trustworthy. In a sense, it is the most transparent and least socially intelligent 'machines' that we trust the most. If you set your microwave for three minutes, you know it will cook for three minutes. One can hardly count on, say, a spouse to follow instructions so precisely. Presumably, a robot that adopted all the tricks of the con artist would be the most inspiring of trust but the least deserving of it.

The degree of *cooperation* human beings are willing to give to a robot [28] is highly dependent on exogenous factors. For example, a person's cooperation may depend on the robot's power to perform a necessary task irrespective of its social ability. *Reciprocity* is closely related to cooperation. To some extent, reciprocity presupposes shared goals, values, group affiliation, friendship, or some other basis for cooperation. If the situation favors reciprocity, it is a useful benchmark. Nevertheless, one must distinguish the *capacity* for reciprocity from the *intention* or *motivation*. They may result in the same behavior, although cognitively and ethically they are different. A robot that has the capacity to reciprocate but lacks the intention may, in certain circumstances, be perceived as being antisocial or even blameworthy, but this does not necessarily make it less of a person. Reciprocity varies depending on the kind of relationship. We might expect to see it more in cooperative relationships among peers than in parent-child or competitive relationships.

IV. Concluding remarks

An ability to form and maintain long-term relationships is emblematic of personhood. Given that we have tried to avoid a species-centric view of personhood, to what extent should robots be built in the image of human beings?

One reason to build such android robots is to study *human* interaction, because, if we consider all the kinds of experimental apparatuses we could build, androids would be the most likely to elicit the kinds of responses that human beings direct toward each other. A second reason

[4]Holocaust survivors from religious backgrounds often concluded that God had died or gone insane [39].

[5]In that sense, a mother can preserve herself by forfeiting her life so that her child, whom she loves and identifies with, may live.

to build androids is to test cognitive theories about human beings. It is advantageous to implement theories of human cognitive function in androids rather than less human-looking robots, because we are highly sensitive to the violation of human norms in humanlike forms. This sensitivity provides information useful to critiquing cognitive theories and is one aspect of what Masahiro Mori referred to as the *uncanny valley* [41]. These two reasons are at the heart of the new field of *android science* [4], and they are motivated by a desire for a deeper understanding of human beings. However, they are not meant to be broad justifications for building human-looking robots.

If our goal is to build robots that can turn themselves into people, there is a very practical reason for making these robots look human. Human beings are the most paradigmatic examples of persons that we know of. Therefore, the focus should be on evaluating robots in terms of their relationships with human beings. Human beings are most adept at interacting with entities that look and act human. Our brains have co-evolved with our expressive bodies and faces and have been honed by experience to understand human feelings and intentions. Friendly extroverts who cannot fully animate their faces because of Parkinson's disease or Moebius Syndrome can be mistaken for sullen introverts and suffer social isolation [42]. So rather than cripple robots with a form that human beings are less adept at interpreting, it makes sense for robot engineers to leverage on the human form by building very humanlike androids.

REFERENCES

[1] R. Ryder, *Animal revolution: Changing attitudes towards speciesism*. Oxford: Blackwell, 1989.

[2] T. Regan, *The case for animal rights*. Berkeley, California: University of California Press, 1983.

[3] K. MacDorman and H. Ishiguro, "Toward social mechanisms of android science: A CogSci 2005 workshop," *Interaction Studies*, vol. 7, no. 2, 2006.

[4] ——, "The uncanny advantage of using androids in social and cognitive science research," *Interaction Studies*, vol. 7, no. 3, 2006.

[5] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, pp. 433–460, 1950.

[6] J. Weizenbaum, "Eliza: A computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[7] S. Harnad, "Lost in the hermeneutic hall of mirrors: An invited commentary on Michael Dyer's Minds, machines, Searle and Harnad," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 2, pp. 321–327, 1990.

[8] N. Block, "Psychologism and behaviorism," *Philosophical Review*, vol. 90, pp. 5–43, 1981.

[9] R. M. French, "Subcognition and the limits of the Turing Test," *Mind*, vol. 99, no. 393, pp. 53–65, 1990.

[10] S. Harnad, "The Turing Test is not a trick: Turing indistinguishability is a scientific criterion," *SIGART Bulletin*, vol. 3, no. 4, pp. 9–10, 1992.

[11] ——, "Minds, machines and Turing: The indistinguishability of indistinguishables," *Journal of Logic, Language, and Information*, vol. 9, no. 4, pp. 425–445, 2000.

[12] R. M. French, "The Turing Test: The first fifty years," *Trends in Cognitive Sciences*, vol. 4, no. 3, pp. 115–121, 2000.

[13] D. Ross, "Emotions as strategic signals," *Rationality and Society*, vol. 16, no. 3, pp. 251–286, 2004.

[14] S. Cowley, "Contextualizing bodies: How human responsiveness constrains distributed cognition," *Language Sciences*, vol. 26, no. 6, pp. 565–591, 2004.

[15] ——, "Beyond symbols: How interaction enslaves distributed cognition," 2005.

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

[17] W. Christensen and M. Bickhard, "The process dynamics of normative function," *Monist*, vol. 85, pp. 3–28, 2002.

[18] R. Hinde, *Individuals, Relationships, and Culture*. Cambridge: Cambridge University Press, 1987.

[19] L. Wittgenstein, *Philosophical investigations*. Oxford: Blackwell, 1953.

[20] J. Shotter, "'Now I can go on': Wittgenstein and our embodied embeddedness in the 'hurly-burly' of life," vol. 19, pp. 385–407, 1996.

[21] E. Goffman, *Forms of Talk*. Oxford: Blackwell, 1981.

[22] ——, "The interaction order," *American Sociological Review*, vol. 48, pp. 1–17, 1983.

[23] K. Dautenhahn and I. Werry, "A quantitative technique for analysing robot-human interactions," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[24] M. Goodrich and D. Olsen, "Seven principles of efficient human robot interaction," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 2003, pp. 3943–3948.

[25] D. Olsen and M. Goodrich, "Metrics for evaluating human-robot interactions," in *Proceedings of the NIST Performance Metrics for Intelligent Systems Workshop*, 2003.

[26] P. Kahn, *The human relationship with nature: Development and culture*. Cambridge, Massachusetts: The MIT Press, 1999.

[27] P. Kahn, H. Ishiguro, B. Friedman, and T. Kanda, "What is a human? Toward psychological benchmarks in the field of human-robot interaction," in *RO-MAN 2006 special session: Psychological benchmarks of human-robot interaction*, 2006.

[28] J. Goetz and S. Kiesler, "Cooperation with a robotic assistant," in *Proceedings of computer-human interaction*, 2002.

[29] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *Proceedings of the AAAI Fall Symposium on Emotional and Intelligent II: The Tangled Knot of Social Cognition*, 2001.

[30] J. Schulte, C. Rosenberg, and S. Thrun, "Spontaneous, short-term interaction with mobile robots in public places," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 1999.

[31] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Coodrich, "Common metrics of human-robot interaction," in *Proceedings of Human-Robot Interaction*, 2006.

[32] K. Dautenhahn, I. Werry, J. Rae, P. Dickerson, P. Stribling, and B. Odgen, "Robotic playmates: Analysing interactive competencies of children with autism playing with a mobile robot," in *Socially intelligent agents: Creating relationships with computers and robots*, K. Dautenhahn, A. Bond, L. Canamero, and B. Edmonds, Eds. Kluwer, 2002.

[33] B. Reeves and C. Nass, *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press (CSLI), 1996.

[34] J. Lee and K. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, pp. 50–80, 2004.

[35] J. Baldwin, *Social and ethical interpretations in mental development: A study in social psychology*. New York: Arno, 1897/1973.

[36] S. Cowley, "Distributed cognition at three months: Mother-infant dyads in kwaZulu Natal," *Alternation*, vol. 10, no. 2, pp. 229–257, 2003.

[37] T. Pyszczynski, J. Greenberg, and S. Solomon, "A dual-process model of defense against conscious and unconscious death-related thoughts: An extension of terror management theory," *Psychological Review*, vol. 106, no. 4, pp. 835–845, 1999.

[38] S. Solomon, J. Greenberg, and T. Pyszczynski, "Tales from the crypt: The role of death in life," *Zygon: Journal of Religion and Science*, vol. 33, pp. 9–43, 1998.

[39] E. Wiesel, *Night*. New York: Bantam Books, 1960.

[40] B. Bettelheim, *The children of the dream*. New York: Macmillan, 1969.

[41] M. Mori, "Bukimi no tani [the uncanny valley]," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[42] J. Cole, "Empathy needs a face," *Journal of Consciousness Studies*, vol. 8, no. 5–7, pp. 51–68, 2001.