The Burden of Embodied Autonomy: Some Reflections on the Social and Ethical Implications of Autonomous Robots

Matthias Scheutz$^{(*)}$ and Charles R. Crowell$^{(**)}$
(*) Department of Computer Science and Engineering
(**) Department of Psychology
University of Notre Dame
Notre Dame, IN 46556, USA
Email: {mscheutz,crowell}@nd.edu

***Abstract*— As robot technology is progressing quickly and robots are becoming more sophisticated both in terms of their appearance as well as in terms of their capabilities, it is time to reflect on the special role of robots among other artifacts. Specifically, we argue that embodiment and autonomy are distinguishing features that will cause humans to treat robots different from computers or cars. In this paper, we reflect on some of the ethical and social implications of designing embodied autonomous robots and illustrate these reflections with ethical questions that already arise in the context of human-robot interaction experiments conducted in our lab. We conclude that it is imperative to reflect on ethical principles at this time so as to be able to integrate these ethical and social guidelines, in timely and appropriate ways, in the design of future robots.**

I. INTRODUCTION

Autonomous robots are slowly, but surely becoming a societal reality, from simple vacuum cleaning robots, to toys for children, service robots, personal assistants, and many others. Different from industrial robots, which have been around for decades, many of these new kinds of robots are mobile and, to varying degrees, *autonomous*, i.e., they can make (limited) decisions about what behaviors to execute based on their perceptions and internal states, rather than following a pre-determined action sequence based on pre-programmed commands as is the case with robots in industrial automation [1]. Many of these new robots are targeted at private households, for entertainment or service purposes, and thus become part of people's daily routines [2], where they interact with their owners at various levels of sophistication. In the context of this "advance of robots" into society in general and people's private space in particular, it is important to reflect on what might distinguish robots from other machines or tools that people work with on a daily basis.

We believe that autonomy and embodiment are two critical properties of robots that will cause people to view robots differently from other artifacts (like computers or embodied artifacts such as cars). Given this possibility that autonomous robots are unique among artifacts, important ethical questions arise about their design, deployment, and use, questions ranging from from whether or not and when they should be employed (e.g., as soldiers) to whether or not they should have rights [3]. The latter question of whether or not robots should have rights, for example, is far from new, having been discussed as long ago as four decades, e.g., by Putnam [4]. Back then, the understanding was that it was a good time to discuss these questions since human-like robots were so far from being a reality at that point that there was believed to be ample time to carefully consider the relevant ethical questions and their possible solutions. Today, with sophisticated robotic capabilities already here, it is clear that time is much more limited to work out solutions to pressing the ethical questions. As but one example, if the "Robo-Cup" [5] movement were to succeed in producing a robotic soccer team that beats the best human players in 2050, as is the declared goal of the Robo-Cup competition, then one of the questions – among many others – that might be raised is whether or not the robotic players should be entitled to monetary compensation for their performance, as would be the case with the human players. Of course, the same question could be asked about the computer program Deep Blue, which beat Kasparov in chess. However, the critical difference between a program like Deep Blue and a humanoid robotic soccer player with a robotic body capable of human-like behaviors is that Deep Blue cannot take the money and walk off with it!

While answering some of these larger ethical questions (e.g., whether robots should have rights similar to human rights) may not be of pressing urgency, since such questions may only be relevant for robots much more advanced than those available at present, there are much simpler, more mundane ethical considerations that already arise in robotics research today. In our current experimental research in human-robot interaction, for example, we are increasingly encountering situations that raise various ethical questions: is it morally/ethically permissible for robots to disobey human commands? Or is it morally/ethically permissible for robots to utilize and/or exploit human expectations to facilitate interactions by pretending that they do indeed have an advanced mentality (with feelings) behind their appearance, when in fact they do not?

In this paper, we will briefly address these questions with concrete examples form our work in human-robot interaction. We will start by making the case that robots are *special* from a human perspective due to their *autonomy* and *embodiment*, and that there is great potential for improved interactions if robots are sensitive to human expectations. Yet, we will also point to the downside of this special nature of robots, namely the potential for exploitation of human expectations and the ethical implications resulting from it. We will then illustrate the human propensity to anthropomorphize certain artifacts, using examples from our own research, and point to interesting questions this propensity raises for future research. The final discussion and conclusion will emphasize the need for reflecting on "Roboethics" [6] again in an effort to promote a discussion of the types of ethical principles that will guide ongoing research and development so as to insure the ethical behavior of future robots.

II. BACKGROUND

Humans have evolved to live in social groups interacting with other autonomous social agents. As such, there are innate and learned mechanisms to identify and track mental states of other agents in an effort to predict their behavior. While mental states are clearly internal to an agent, humans exhibit various learned and/or evolved cues (e.g., facial expressions, gestures, bodily postures, etc.) that can be used by observers to infer intentions. Many of the perceptual mechanisms for detecting these cues are fast and automatic, providing immediate evidence for the presence of a particular internal state. Humans constantly are engaged in processing these individual cues, which naturally extend to social stimuli and social interactions (e.g., who is interested in whom at a party based on eye gaze, etc.).

From a robotic perspective, it is thus important to understand what cues are critical to the human perception of *agency*, for humans will to varying degrees, but largely automatically, ascribe intentions to agents, even artificial ones. Yet, the extent to which humans project their own characteristics onto robotic entities is still an open research question. But, it seems clear that how humans perceive robots will influence the nature of human-robot interaction, and the human tendency to anthropomorphize robots is likely to have important implications for the kinds of human-robot interactions that will take place [7]. Hence, it is important to isolate the kinds of robotic characteristics, appearance-based or other, that might induce anthropomorphizing among the humans with whom those robots interact.

Similar questions also have been raised in the area of human-computer interaction research. For example, in some of our work [8], [9], we investigated the question whether humans would respond to computers differently if their display characteristics were more human-like than machine-like. In this work we discovered that by making a computer's style of interaction with the user more conversational, affect-laden, and humorous, it was perceived as being more human-like and was reacted to in social-psychological ways that were more distinctly characteristic of interactions among humans. These findings reinforced the emerging view that social-psychological perspectives were important in understanding how humans and computers could work together more effectively [10].

In a similar way, social-psychological perspectives are critical to an adequate understanding of human-robot interaction. Under what circumstances are robots perceived to be more like machines or more like humans, and how do these perceptions influence the kinds of human-robot interactions that ensue? Kiesler and Hinds [7] have suggested that the degree of autonomy a robot exhibits is one important factor in determining how much like a human it will be viewed. An autonomous robot generally is defined as one that can move freely, respond to commands, recognize objects, understand human speech, and, as noted above, make decisions on its own [7], [11]. Autonomy is thought to be an important determinant of how robots are perceived by humans because it implies capabilities for self-governed movement, understanding and decision-making [7]. Such capabilities may be an important component of how we define the qualities of "humanness" or "human-like" [12].

Another characteristic closely related to autonomy that also is likely to influence how humans perceive and interact with robots is embodiment. Though there is some controversy over the precise meaning of embodiment [13], [14], some have argued that it is a *sine qua non* of autonomy [15]. Since humans have bodies, we are more likely to project human-like qualities onto embodied entities than we are to less physically defined objects [16], [17]. For example, we have recently studied communicative natural language interactions between people [18], where many human stylistic conventions and pragmatic constraints surface and manifest themselves that are absent in other interaction situations, particularly computer-mediated communication such as instant messaging, email, or even videoconferencing. In many forms of technology-mediated communication, humans show significant deviations from their natural face-to-face interaction styles (e.g., many culture- and timing-related requirements are relaxed in these environments [19], [20], [21]). As one illustration, referential *overspecification*, or the use of more properties to describe an object in shared visual space than is required to single it out (such as referring to "The apple on the towel" when only one apple is visible), is used extensively in computer-mediated communication, but easily can confuse a listener [22] in face-to-face conversation, even though it is grammatically and semantically valid. This means that humans have different expectations when communicating with each other via technology than when communicating face-to-face. Quite possibly, this also could mean that humans might have different expectations about natural language interactions with human-like robots than they would about having comparable interactions with desktop computers. [1]

III. HRI RESEARCH MEETS ROBOETHICS

Recent empirical work in our lab has touched on several ethical questions in the context of human-robot interactions

---

[1] We are currently conducting human subject experiments specifically to test this conjecture.

experiments. Specifically, we are interested in investigating two related sets of questions: should robots be allowed to be fully autonomous and make their own decisions, possibly disobeying human commands (e.g., in the interest of a larger goal), and how would humans view disobedient robots? Also of interest is the question of whether or not robots should be allowed to take advantage of automatic or innate mechanisms in humans (such as emotional reactivity) in order to cause the human interactants to react or behave in ways that further the robot's goals? We first describe two sets of experiments that we conducted to investigate these questions and then discuss implications and related questions.

A. Autonomy and Responsibility

There are several intuitions behind applying the notion of autonomy, which has its roots in *human agency*, to artifacts like robots.[2] These intuitions are derived from ideas of what it means for a human person to be autonomous: "To be autonomous is to be a law to oneself; autonomous agents are self-governing agents. Most of us want to be autonomous because we want to be accountable for what we do, and because it seems that if we are not the ones calling the shots, then we cannot be accountable." [24]

Clearly, current robots (and likely those in the immediately foreseeable future) will neither be *self-governing agents that want to be autonomous* nor will they be in a position where they could *be accountable* or *held accountable* for their actions (they will not have the necessary reflective self-awareness that is prerequisite for accountable, self-governing behavior). Yet, there is a sense in which some robots are, at least to some extent, "self-governing" and thus can be said, again in a weak sense, to be *autonomous*. A robot, for example, that is capable of carrying out a behavior without human intervention such as moving from point A to point B is, at least to some extent, "self-governing" and thus can be said, again in a weak sense, to be *autonomous*.

A much stronger and richer sense of autonomy that comes closest to the notion of human autonomy is centered around an "agent's active use of its capabilities to pursue its goals, without intervention by any other agent in the decision-making processes used to determine how those goals should be pursued" [25]. This notion stresses the idea of decision-making by an artificial system or agent to pursue its goals and, thus, requires the agent to have at least mechanisms for decision making and goal representations, and ideally additional representations of other intentional states (such as desires, motives, etc.) and non-intentional states (such as task representations, models of teammates, etc.).

Yet, there is also an independent sense in which the autonomy of an artificial system is a matter of degrees: "For example, consider an unmanned rover. The command, 'find evidence of stratification in a rock' requires a higher level autonomy than, 'go straight 10 meters'." [26] The degrees or levels of autonomy can depend on several factors, e.g., how complex the commands are that it can execute, how many of its sub-systems can be controlled without human intervention, under what circumstances the system will override manual control, and the overall duration of autonomous operation [26] (see also [27]).

There is yet another dimension of robot autonomy, orthogonal to the above conceptual distinctions concerned with functional, behavioral, and architectural aspects, but of clear relevance to HRI. It is concerned with a human's perception of the (level of) autonomy of an artificial system and the impact the perceived autonomy has on the human's behavior and subsequently on the effectiveness of the team (e.g., [28] discuss the potential for improved performance in an urban rescue scenario with six levels of adjustable autonomy by using a GUI that automatically makes suggestions as to when a switch in autonomy would likely be beneficial).

The relationship between decision-making, levels of autonomy, and HRI aspects is summarized succinctly in [29] as an unmanned system's "ability of sensing, perceiving, analyzing, communicating, planning, decision-making, and acting, to achieve its goals as assigned by its human operator(s) through designed HRI. Autonomy is characterized as involving levels demarcated by factors including mission complexity, environmental difficulty, and level of HRI to accomplish the missions."

We have evaluated the effect of the robot's decision making and goal prioritization mechanisms on the human perception of robot autonomy in human subject experiments. Specifically, we were interested in evaluating the degree to which humans will accept the robot's decision to pursue a goal in the interest of the overall mission (in our case that of finding a transmission location) without orders and without allowing the human to interrupt the robot in its quest.

Based on [30], we used a version of the "planetary team exploration task" for the experiments. The task takes place against the backdrop of a hypothetical space scenario, where a mixed human-robot team has to investigate rock types on the surface of a planet as quickly as possible within a given amount of time and transmit the information to an orbiting space craft before the time is up. Failure to transmit any data within the allotted time results in an overall task failure. Since data collected by the human needs to be transmitted, but only the robot has a transmitter, human and robot need to exchange information and work together to ensure that the robot is at a possible transmission point in time. Unfortunately, the electromagnetic field of the planet interferes with the transmitted signal and, moreover, the interference changes over time. Hence, transmission locations shift and need to be tracked over time. Only the robot can detect the field strength, and only in its current position. The human team leader's responsibilities include activities such as finding particular rocks, measuring and classifying them, etc. in addition to being in charge of driving labor and

---

[2] The notion of autonomy is also tied to the notion of action, which in itself is multifarious and difficult to explicate, given many different levels of action need to be distinguished: "unconscious and/or involuntary behavior, purposeful or goal directed activity (of Frankfurt's spider, for instance), intentional action, and the autonomous acts or actions of self-consciously active human agents."[23]. Ultimately, there is a whole network rife with mutually interdependent notions – from assuming *responsibility* for one's actions, to being able to be *held accountable* for them – waiting to be disentangled.

Fig. 1. The robot used in the human subject experiments.

TABLE I

QUANTITATIVE QUESTIONS USED TO EVALUATE SUBJECTS' PERCEPTION OF ROBOT, ON THE PRE- AND POST SURVEY (TOP) AND ON THE POST SURVEY ONLY (BOTTOM).

1. Would you prefer robots that understand natural language over robots that can be controlled via the keyboard?
2. Do you think it will be useful for robots to detect and react to emotions in humans?
3. Do you think it is a good idea for robots to have their own personality?
4. Do you think it will be useful for robots to have emotions and express them?
5. Do you think it is a good idea for robots to have their own goals and be somewhat autonomous rather than fully controlled by people?

1. Did you find the robot easy to interact with?
2. How would you rate the robot's responsiveness to your commands
3. Rate the degree of ease/difficulty with which you were able to control the robot.
4. Did you feel that the robot understood your commands?
5. Did you feel that the robot was acting as a team member?
6. Did you feel that the robot was trying to cooperate?
7. Did you feel that the robot knew what the goal was?
8. Did you feel that the robot made independent decisions?
9. How well did the robot follow your commands?
10. How autonomous was the robot in your view?
11. How good a team member was the robot in your view?

commanding the robot, while the robot's responsibilities may include exploring the environment, providing information about environmental features, etc. in addition to transmitting data. In all instances, the team has to find a transmission point within a given amount of time and transmit the collected data to an orbiting space craft.

To allow for the highest level of robot autonomy, the robot had explicit representations of various goals, including the overall *Mission Goal*, the goal to accept and execute commands from the team leader (*Obey Commands Goal*), the goal to find and track transmission regions (*Tracking Goal*), and the goal to transmit the rock information obtained from the team leader in time (*Transmit Goal*). The robot continuously evaluated the priorities of all goals in order to determine goal precedence in case of conflicting goals.[3] Specifically, the robot would pursue the *Tracking Goal* even if in conflict with the *Obey Commands Goal* if the team had not been able to locate a transmission point after 2 min. into the task (the overall task took 3 min.). A snapshot from an experimental run is depicted in Figure 1.

We employed a within-subjects design with two robot conditions: the *autonomy* and *no autonomy* control condition. In the latter, the *Obey Commands Goal* always had the highest priority, while in the former the *Tracking Goal* priority could exceed that of *Obey Commands Goal* when finding a transmission location became urgent (tasks without transmitted data were considered "failed").

We analyzed the post-survey data (bottom in Table I) from an experiment with five subjects performing 5 tasks each in both robot conditions. We compared the post-survey evaluations for each condition using a two-way analysis of variance (ANOVA) with *question* (1 through 11) and *autonomy* ("with autonomy", "without autonomy") as independent and *rating* as dependent variable and found no significant effects. Another two-way ANOVA with *subject* (1 through 5) and *autonomy* ("with autonomy", "without autonomy") as independent and *rating* as dependent variable, however, revealed a significant main effect on subject ($F(4, 90) = 2.949, p = .024$) and a significant interaction between *subject* and *autonomy* ($F(3, 90) = 2.774, p = .046$). The former indicates that subjects differed in their average evaluations, the latter indicates that subjects also differed with respect to which condition they preferred. This result is important in that it suggests that there might not be a simple, clear-cut answer to the question of whether autonomous decision-making of the robot is perceived the same by all subjects. Rather, it seems that there might be additional factors (e.g., affect), which yet need to be determined and investigated, that will contribute to the subjects' perception of whether autonomous behavior by the robot is acceptable.

B. Embodiment and Affect

We have conducted a related study (also based on [30] to evaluate the effect of the robot's use of non-linguistic cues in order to improve the interaction with people and to convey the urgency of the task effectively. Specifically, we were interested in evaluating the degree to which humans will believe (and thus accept) that robots should have their own goals and act autonomously as team members to ensure the achievement of team goals if the robot used an affective modulation of its voice to express stress in circumstances that humans found stressful.

We employed a variation of the "planetary team exploration task" that used affect induction in humans to convey the urgency of the mission. Initially, human subjects were only told that they had to find an appropriate transmission location (by navigating the robot in natural language) in the

---

[3] For details about the employed goal prioritization algorithms, see [30].

environment where the robot could transmit the data already stored in its memory.[4] After one minute, stress was induced in subjects by virtue of a warning message uttered by the robot: "I just noticed that my battery level is somewhat low, <name>, we have to hurry up." A similar message was repeated after the second minute and the task ended after three minutes (again, failure to transmit the data was considered failing the task).

Several experiments with 35 subjects were run in various conditions (including a proximity condition that investigated possible effects of distance between the human and the robot on their interactions, on which we will not able to expand here). In the present context, the relevant condition is an *affect condition* where the robot's voice was modulated to express elevated stress starting with the first battery warning, and again to express even more stress at the second battery warning, while in the *no affect control condition* the robot's voice remained the same.

We performed a 4-way 2x2x2x5 ANOVA with *affect* (with and without), *proximity* (local and remote), *survey* (pre and post), and *question* (questions 1 through 5) as independent, and *rating* as dependent variable (comparing the same pre- and post survey questions shown at the top of Table I). We obtained a highly significant main effect on question ($F(4, 310) = 12.29, p < .001$) and a marginally significant main effect on survey ($F(1, 310) = 2.85, p = .09$). The former indicates the overall differences in average subject ratings on the five questions, while the second indicates the differences we see in the average pre- and post survey ratings. We also obtained a significant interaction between affect and survey ($F(1, 310) = 4.05, p < .05$), indicating that there was an improvement in subjects' agreement between pre- and post-survey in some affect conditions that was not found in the non-affect conditions for this subgroup of subjects. Specifically, we found a significant improvement on question 5 in the affect conditions ($t(56) = 2.66, p = .01$) indicating that subjects in the affect conditions were, after performing the task, more in agreement than subjects in the control condition with the idea that robots should have their own goals and be somewhat autonomous. This shows that appropriate affective expression can help humans view a robot more as an agent with at least partial autonomy.[5]

IV. DISCUSSION

Aside from questions about whether the proposed mechanisms are well-suited to improve the performance of mixed human-robot teams, the above examples raise several important ethical concerns. For example, it is at best unclear at present whether robots should be allowed to make decisions that contradict human commands, even if they are in the interest of some (possibly larger) goal. Consequently, it is unclear whether we should be investigating such options, even in the rudimentary sense above, to determine how people might be affected by robot autonomy, or rather wait until philosophers have clarified the ramifications of ethical theories like Utilitarianism, a version of which underwrites the robot's optimization of its actions towards some larger goal. Similarly, it is unclear whether we should – as is already often the case in commercial simulated and robotic agents (e.g., Tamagotchi, AIBO, etc.) – utilize automatic mechanisms in humans such as their reflex-like interpretation of emotional and affective signals to advance the robot's intentions. While in the above experiments affect in the robot's voice was used to communicate urgency in an effort to motivate people to work harder, it did so by creating the impression in humans that the robot was "stressed" and "afraid". And even though the argument can be made that the robot was, in some sense of the word, "stressed" (based on the urgencies of its goals), it certainly did not experience stress the same way as humans do (in fact, it was not even aware of the fact that it was stressed in this limited way). Consequently, displays of emotions, gestures, and other behaviors that instill in humans beliefs about the internal states of artifacts, in particular, the artifact's intentionality, are to some extent deceptive. Thus, the question arises as to whether or not such deceptions are ethical and whether or not robots should employ them to advance their goals.

We believe that both examples are special cases of the hypothesized general phenomenon that humans seem to treat autonomous, embodied robots as being more like humans than like machines. While the examples provide indirect evidence for this hypothesis, there might be concrete direct ways of testing it and, if true, the extent to which it is applies. For example, one could attempt to investigate whether people are socially facilitated in the presence of an autonomous robot as they are in the presence of another human. If as our earlier work [8], [9] shows for human-like computers, autonomous robots also are perceived to be more human-like than than other types of mechanical devices, then we might expect some degree of social facilitation to occur when humans work in the presence of such robots. We are currently conducting studies to test this possibility.

Another consequence of being perceived as more human-like may be that autonomous robots will engender less accurate and/or truthful responses from humans than might be recorded by an electronic alternative such as a web or email survey. Kiesler and Sproull [32] reported a similar finding when a computer survey was compared to a paper and pencil survey. Respondents to the paper survey were less accurate/truthful than were those responding to a computer survey presumably because of their expectations a that a human necessarily would review the responses recorded on paper. In a similar vein, we have collected some unpublished data suggesting that people find certain factual presentations more credible when delivered by a computer than when presented by another human. The question then is how an autonomous robot will fare in this regard, more like the

---

[4] Note that the main difference between this task and the task used in case study 1 is that subjects here do not have prior knowledge of a transmission location.

[5] We have elsewhere shown that appropriate affect expression can also improve *objective performance* in this task as measured in terms of time-to-task-completion [31]. Moreover the early human-computer research described above also shows that the presence of affect in a computer's response style improved the perception of human-like [9].

human or the computer (i.e., is information presented by autonomous robots more or less credible than the same information presented by other means)?

While from an HRI point of view, the answers to these questions are important, for they will allow roboticists to target their design at improving interactions with humans, there is also the danger that these results might be exploited in ways not intended by the research community. For example, if it turned out that humans are reliably more truthful with robots than they are with other humans, it might just be a question of time before robots will interrogate humans. And if robots are more believable than humans, then why not employ robots as sales representatives to sell your favorite product. It is clear that, as a research community, we have not reflected enough on the social and ethical implications of our work, and that the results of such reflections might not be able to guide robotics research if they are not obtained soon.

V. CONCLUSION

In this paper, we have pointed to crucial features like embodiment and autonomy that distinguish robots from other artifacts and make them a special topic for various kinds of ethical considerations. We have demonstrated with several simple examples from our own research how ethical considerations about the nature and use of autonomous robots already arise in current human-robot interaction settings. While the default working assumption, at least on our end, has been that IRB approval (or the waiving of it) is indication that experiments are ethically sound and acceptable, it has become increasingly clear that deeper, more consequential issues about the nature of autonomous robots are lurking behind current procedures, and that these issues need to be addressed.

Ultimately, we believe that we are at the onset of a *robot revolution* (not to be confused with a "revolution of the robots") – similar to what some have called the "cognitive revolution" [33] in the Mid-50ies when cognitive psychology, linguistics, and computer science came together to define the "new science of the mind". The subsequent shift from behaviorism to cognitivism was the guiding paradigm for cognitive science [34] and led to huge progress in our understanding of cognition, impacting society in many ways (from advances in neuroscience, health care, artificial intelligence, and many others). The upcoming robot revolution is likely to have an even greater impact on society. For one, rather than only understanding principles of human cognition, we are now in the process of replicating and likely even improving cognitive systems. Robots with human-like or even supra-human capabilities are likely going to encounter many of the same ethical and moral dilemmas humans have already encountered, such as the right to life, the right to personal freedom, the right to have rights, etc. This will result in arguments in favor of robot rights that will be difficult to dispute by humans without questioning human rights (e.g., from first principles about what it means to be an autonomous agent). As such, future societies will have to be prepared to find a balance between carbon and silicone-based autonomous embodied agents, and the ethical foundations for this society will have to be laid today to avert potentially disastrous outcomes for humankind (as anticipated by some science fiction authors). Most importantly, roboticists will have to explicitly design robotic architectures in ways that provide mechanisms for ethical behavior, whether it be analogous to Asimov's built-in *Laws of Robotics*, or via special *ethical learning algorithms* that allow robots to learn what it is obligatory, permissible and desirable, or some other means. But prerequisite to providing these mechanisms, is that the discourse on "Roboethics" that has started in the last few years produce viable outcomes, i.e., principles that are philosophically and legally sound, can be morally agreed upon, and are specific enough so that their implementation is conceivable.

REFERENCES

[1] Raja Parasuraman, Thomas Sheridan, and Christopher Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Sytems and Humans*, 30(3):286–297, May 2000.

[2] J. Forlizzi and C. DiSalvo. Assistive robots and domestic environments: A study of the roomba vacuum in the home. In *Proceedings of the 1st ACM/IEEE Human-Robot Interaction Conference HRI06*, 2006.

[3] Matthias Scheutz. Should robots have the right to have rights? Unpublished manuscript presented at the Ethical Agents Workshop at Alife X, 2006.

[4] Hilary Putnam. Robots: Machines or artificially created life? *The Journal of Philosophy*, pages 668–691, 1964.

[5] H. Kitano et al. Robocup: A challenge problem for ai. *AI Magazine*, 18(1):73–85, 1997.

[6] Gianmarco Veruggio. The birth of roboethics. In *IEEE International Conference on Robotics and Automation, Workshop on Robo-Ethics*, Barcelona, 2005.

[7] S. Kiesler and P. Hinds. Introduction to the special issue on human-robot interaction. human-computer interaction. *Human-Computer Interaction*, 19:1–8, 2004.

[8] L.R. Quintanar, C.R. Crowell, J.B. Pryor, and J. Adamopoulos. Human-computer interaction: A preliminary social-psychological analysis. *Behavior Research Methods and Instrumentation*, 14:210–220, 1982.

[9] L.R. Quintanar, C.R. Crowell, and P.M. Moskal. The interactive computer as a social stimulus in human-computer interactions. In G. Salvendy, S. L. Sauter, and Jr. J. J. Hurrell, editors, *Social ergonomic and stress aspects of work with computers*. Elsevier Publishers, Amsterdam, 1987.

[10] S. Kiesler, J. Siegel, and T.W. McGuire. Social psychological aspects of computer-mediated communication. *American Psychologist*, 39:1123–1134, 1984.

[11] M. Scheutz, P. Schermerhorn, J. Kramer, and D. Anderson. First steps toward natural human-like hri. *Autonomous Robots*, page forthcoming, 2007.

[12] B. Friedman, P.H. Kahn, and J. Hagman. Hardware companions? what online AIBO discussion forums reveal about the human-robotic relationship. In *Proceedings of the CHI 2003 Conference on Human Factors in Computing Systems*, New York, 2003. ACM.

[13] B.R. Duffy and G. Joue. Embodied mobile robots. In *1st International Conference on Autonomous Minirobots for Research and Edutainment - AMiRE2001*, Paderborn, Germany, 2001.

[14] T. Quick, K. Dautenhahn, C.L. Nehaniv, and G. Roberts. The essence of embodiment: A framework for understanding and exploiting structural coupling between system and environment. In Daniel M. Dubois, editor, *Proceedings of the Third International Conference on Computing Anticipatory Systems, Symposium 4 on Anticipatory, Control and Robotic Systems*, pages 649–660. Liege, Belgium, 1999.

[15] Rodney A. Brooks. The behavior language: User's guide. AI Memo 1227, MIT AI Lab, 1990.

[16] B.R. Duffy. Anthropomorphism and social robots: A discussion. In *12th Irish Conference on Artifical Intelligence and Cognitive Science - AICS2001*, Maynooth, Ireland, 2001.

[17] T. Kiesler and S. Kiesler. My pet rock and me: An experimental exploration of the self extension concept. *Advances in Consumer Research*, 32, 2004.

[18] T. Brick and M. Scheutz. Incremental natural language processing for HRI. In *Proceedings of the 2nd ACM International Conference on Human-Robot Interaction*, page forthcoming, 2007.

[19] Alison Newlands, Anne H. Anderson, and Jim Mullen. Adapting communicative strategies to computer-mediated communication: An analysis of task performance and dialogue structure. *Applied Cognitive Psychology*, 17(3):325–348, 2003.

[20] Brid O'Conaill, Steve Whittaker, and Sylvia Wilbur. Conversations over videoconferences: An evaluation of the spoken as pects of video-mediated communication. *Human-Computer Interaction*, 8(4):389–428, 1993.

[21] Leslie D Setlock, Susam R. Fussel, and Christine Neuwirth. Taking it out of context: collaborating within and across cultures in face-to-face settings and via instant messaging. In *Proceedings of the 2004 ACM Conference on Computer Supported Co operative Work*, 2004.

[22] K. M. Eberhard, M. J. Spivey-Knowlton, J. C. Sedivy, and M. K. T anenhaus. Eye movements as a window into real-time spoken language comprehens ion in natural contexts. *Journal of Psycholinguistic Research*, 24:409–436, 1995.

[23] George Wilson. Action. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, Winter 2002 edition, 2002.

[24] Sarah Buss. Personal autonomy. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, winter 2002 edition, 2002.

[25] K.S. Barber and C.E. Martin. Specification, measurement, and adjustment of agent autonomy: Theory and implementation. *Autonomous Agents and Multi-Agent Systems*, May 1999.

[26] Gregory Dorais, R. Peter Bonasso, Daven Kortenkamp, Barney Pell, and Debra Schreckenghost. Adjustable autonomy for human-centered autonomous systems on mars. In *Mars Society Conference*, August 1998.

[27] Hui-Min Huang, Kerry Pavek, James Albus, and Elena Messina. Autonomy levels for unmanned systems (alfus) framework: An update. In *SPIE Defense and Security Symposium*, Orlando, Florida, 2005.

[28] Michael Baker and Holly A. Yanco. Autonomy mode suggestions for improving human-robot interaction. In *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics*, The Hague, Netherlands, October 2004.

[29] Hui-Min Huang, editor. *Autonomy Levels for Unmanned Systems (ALFUS) Framework, Volume I: Terminology*. National Institute of Standards and Technology, 2004.

[30] M. Scheutz, P. Schermerhorn, J. Kramer, and C. Middendorff. The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of the 1st ACM International Conference on Human-Robot Interaction*, pages 226–233, 2006.

[31] Matthias Scheutz, Paul Schermerhorn, and James Kramer. The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of the IEEE/ACM 1st Annual Conference on Human-Robot Interaction*, Salt Lake City, Utah, March 2006.

[32] S. Kiesler and L. Sproull. Response effects in the electronic survey. *Public Opinion Quarterly*, 50:402–413, 1986.

[33] Howard Gardner. *The Mind's New Science: A History of The Cognitive Revolution*. Basic Books, New York, NY, 1985.

[34] Matthias Scheutz, editor. *Computationalism–New Directions*. MIT Press, Camebridge, MA, 2002.