

# John Benjamins Publishing Company



This is a contribution from *Interaction Studies* 8:3  
© 2007. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).

Please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website: [www.benjamins.com](http://www.benjamins.com)

Tables of Contents, abstracts and guidelines are available at [www.benjamins.com](http://www.benjamins.com)

# Working with a robot

## Exploring relationship potential in human–robot systems

Debra Bernstein, Kevin Crowley and Illah Nourbakhsh  
Learning Research and Development Center, University of Pittsburgh /  
The Robotics Institute, Carnegie Mellon University

Research on human–robot interaction has often ignored the human cognitive changes that might occur when humans and robots work together to solve problems. Facilitating human–robot collaboration will require understanding how the collaboration functions system-wide. We present detailed examples drawn from a study of children and an autonomous rover, and examine how children’s beliefs can guide the way they interact with and learn about the robot. Our data suggest that better collaboration might require that robots be designed to maximize their relationship potential with specific users.

**Keywords:** human–robot interaction, educational robotics, robot design

The cognitive sciences have a long history of studying human tool use. Investigations have focused on the way experts engage with the tools in their work environment (Hutchins, 1995), or the way students engage with tools in a learning environment (Van Lehn et al., 2005), to name but a few. While specific interaction goals may differ, the general purpose of the tool remains the same in each situation: to augment the users’ cognitive experience. Research has generally focused on the cognitive changes that accompany tool use.

Robots, particularly those designed for the personal service sector, often support users both cognitively and physically, but most evaluations of human–robot systems fail to investigate the cognitive changes that accompany robot interactions (Stubbs, Bernstein, Crowley & Nourbakhsh, 2006). The difficulty of identifying appropriate cognitive markers for successful human–robot interactions may be one reason why this rarely occurs.

In their paper entitled, “What is a human? — Toward psychological benchmarks in the field of human–robot interaction,” Kahn et al. (2006, 2007) suggest

a potential solution. These authors argue that appropriate cognitive markers for successful human–robot interactions are those that indicate the human user is responding to the robot in humanlike ways. Kahn et al. have identified nine potential benchmarks “that capture conceptually fundamental aspects of human life” (2006, p. 365), and would thus indicate that the user is thinking about the robot as humanlike and responding in kind. The benchmarks are autonomy, imitation, intrinsic moral value, moral accountability, privacy, reciprocity, conventionality, creativity, and authenticity of relation. These benchmarks are psychological in the sense that they measure what the human believes about the robot instead of the robot’s technical capabilities.

The idea of employing benchmarks to measure users’ psychological beliefs about a robot represents an important step forward in the study of human–robot interaction. Psychological benchmarks bring the “human” back into focus by emphasizing the extent to which the human’s beliefs impact successful collaboration. However, we can ask whether the benchmarks previously described represent the only, or even the most appropriate, beliefs to bring into focus. What *should* a human believe about a robot to work with it more effectively?

We take a developmental perspective on this question. Our perspective is developmental in two senses. First, using examples from a study of child–robot interaction, we point out that users, even those who are novice in an absolute sense, already hold established and complex beliefs about robots. These beliefs can guide how the human interacts with robots and may need to be changed if they encourage non-adaptive behavior. Belief change is a developmental process in which old and new knowledge often struggle against each other before the new beliefs can take hold. Which brings us to the second sense in which our perspective is developmental. Although we present snapshots of initial child–robot encounters, we argue that the key to long-term successful interaction hinges on describing, understanding, and eventually designing to support the way that human beliefs about robots change in the context of ongoing collaborative relationships.

### *Humanlike robots vs. robots that work with humans*

What is the best way to seed successful human–robot interactions? One way is to create robots with human characteristics. The idea of creating humanlike machines presents a fantastic but extremely difficult challenge to robotics. Part of the challenge lies in simply identifying those characteristics that capture fundamental humanness. MacDorman and Cowley (2006) point out that some of the characteristics previously identified as “fundamentally human” may reflect a set of culturally-specific moral values rather than a universal human code. This objection underscores the difficulty of identifying benchmarks that capture the essence

of all human beings regardless of their socio-cultural context. Stated another way, “there is no such thing as a generic human being that can be used in a standardized benchmark” (MacDorman & Cowley, p. 379).

We would also argue that human likeness is not a pre-requisite for a successful human–robot relationship. While there is certainly evidence suggesting that people have successful interactions with humanoid robots (Kanda, Hirano, Eaton & Ishiguro, 2004; Minato, Shimada, Ishiguro & Itakura, 2004), some of these research paradigms exclude nonhumanoid controls, so it is difficult to conclude that the humanness of the robot was essential to the successful interaction. But perhaps the best evidence comes from the numerous successful social robots that neither resemble nor imitate people (Fong, Nourbakhsh & Dautenhahn, 2002); for example, the robotic seal *Paro* has been successful in increasing social and physiological functioning in the elderly (Wada & Shibata, 2006). Other examples include the robotic dog *AIBO* (Kahn, Freier, Friedman, Severson & Feldman, 2004) and even the *Roomba* robotic vacuum cleaner (Forlizzi & DiSalvo, 2006). There is certainly variation in the types of relationships users can have with these technologies. The vacuum robot, for example, was sometimes viewed as a worker with household responsibilities, while the robot dog was seen as a playmate and companion. However, in each case, users developed beliefs appropriate to the robot’s perceived role.

As these examples indicate, people are able to engage with intelligent tools that do not exhibit human characteristics. With experience people can become adept at communicating with technology in distinctly nonhuman ways. As creators of intelligent tutoring systems have recently learned, attempts to make technology more humanlike are not always necessary or successful.

Intelligent tutoring systems (ITS) are artificial intelligence (AI) systems that construct cognitive models of students working in scaffolded problem-solving environments. The tutors monitor student progress and make instructional suggestions based on their diagnosis of the mismatch between the student’s current model and the optimal instructional model. First generation tutors taught using a static instructional model. Second generation tutors added student modeling that guided instruction based on student performance. It was believed that third generation tutors, which would use spoken natural language to interact with students, would make the systems more powerful because they would go beyond the limits of human–computer interaction and bring the tutors into the realm of human–human tutoring. The construction of third generation tutors was a massive technological enterprise and one that largely failed to produce improvements in student learning outcomes (e.g., Litman et al., 2005).

As the ITS example shows, increasing parity between human–human interactions and human–technology interactions does not necessarily result in more

efficient or successful systems. The positive impact of ITS was largely because of the task analysis and the front-end research that worked out instructional pathways for different content areas and then made the programs responsive to user input (see, for example, VanLehn et al., 2005). In effect, intelligent tutoring systems are efficient learning environments because both the human and machine put in the effort required to build a working relationship and achieve a common, narrowly defined goal.

In the ITS example above, the addition of a humanlike characteristic, spoken natural language, failed to enhance the human–technology relationship. Rather, it was the efforts of both the user and the technology to move towards a common understanding of how to work together that made the relationship valuable. A potentially fruitful role of psychological benchmarks is in measuring that movement towards successful collaboration. We call this concept *relationship potential*, which describes the likelihood that the human and robot will build a successful collaborative relationship. A successful relationship will require the human to develop a set of beliefs about the robot that aid collaboration and will require the robot to clearly communicate capabilities relevant to the collaboration.

### *The importance of autonomy for collaborative success*

As we think about the ways in which humans and robots can profitably work together, it may be useful to consider some of the findings that have emerged from the human–human collaboration literature. For example, Barron (2000) suggests that characteristics such as mutuality of exchanges, joint attention, and shared task goals are common to successful collaborative groups. Good collaboration is not just about doing more work faster. In studies of real world scientific projects, the most productive and sustained collaborations happen among scientists who realize they have to collaborate to solve certain problems (Schunn, Crowley, & Okada, 2005). The unique skills and perspectives of the different collaborators is what enables the successful solution. Human collaborations have a high relationship potential when people share the same goals but have unique roles, when they can learn how to communicate effectively about the problem and solution spaces, when they come to respect and trust their collaborator's responses, and when they begin to enjoy spending time working with their partners. In other words, good collaborations develop from increasingly coordinated and sophisticated exchanges between autonomous agents. The collaborators maintain their autonomy, but have become a shared system of action. Understanding why they are successful is not about reducing their joint work to the contributions of each. It is about understanding how they respond and change as a system.

Sebastian Thrun has argued, “human–robot interaction cannot be studied without consideration of a robot’s degree of autonomy, because it is a determining factor with regards to the tasks a robot can perform, and the level at which the interaction takes place” (2004, p. 14). Kahn et al. (2007) have suggested that autonomy is an important benchmark for predicting the success of human–robot collaborations, in the sense that a human comes to believe a robot is an autonomous social actor. These ways of discussing autonomy presuppose that autonomy is either a technical or psychological characteristic of the robot. Guided by the human–human collaboration literature, we argue for a broader notion of autonomy that characterizes the human–robot system and focuses analysis on both the roles and abilities of the individual actors and the negotiation that occurs when they are working together. Others have evaluated human–robot collaborations at this system-wide level (see Fong, Nourbakhsh, & Dautenhahn, 2002), and we agree with the approach. If we are going to treat humans and robots as legitimate collaborators, they deserve to be evaluated as a collaborative unit.

An example may help to further exemplify the approach. In their work on human users engaging with Roomba in their homes, Forlizzi and DiSalvo (2006; DiSalvo, 2006) describe several users who rearrange the furniture in particular rooms to allow the Roomba to clean as efficiently as possible. For example, one user happily arranged the furniture to create a barrier around a particularly dirty area so that the Roomba could clean that area with maximum efficiency. The point here is not that the Roomba could be an efficient cleaner by itself, but that the human user and the Roomba could work together to maximize the capacity of the system. It is of limited use for the user to think of the Roomba as a truly autonomous actor. Rather, it is more helpful for the user to understand the robot’s autonomous capabilities well enough to use those capabilities to their maximum potential, thus increasing the relationship potential of the system.

In this paper we begin to develop the idea of relationship potential by examining snapshots of initial encounters between children and an autonomous robot. These data will describe how children begin to build a relationship with the robot through their explorations and shared experiences. We propose that such explorations help the user understand what kind of relationship is possible with a particular robot and the type of relationship the user should have with the robot for the human–robot system to accomplish its goals. Our examples focus primarily on the children’s beliefs about the robot and how they responded to its actions. In the discussion, we focus on the complementary question of how robots might be designed to encourage more effective responses from their potential human collaborators.

## Method

Sixty children between the ages of 4 and 7 were invited to interact with the personal exploration rover (PER) at a children's museum. (The PER was located in a quiet room.) These children were initially recruited to participate in a study of their beliefs about intelligence and technology (see Bernstein & Crowley, under review). Afterwards, all but one child accepted our invitation to interact with the PER. These additional data, which are not reported in the prior article, allow us to explore (1) how children gain information about a robot's autonomous capabilities and (2) how their beliefs shape their behavior towards the robot. In this paper, we frame this exploration around the issue of psychological benchmarks.

Children's interactions with the PER were structured only by the goal they were given — to get the PER to move to a rock on the opposite side of the room. There was no control interface available. Rather, they were encouraged to interact with the rover in any way they wanted to accomplish the goal.

The PER was initially designed as part of a museum exhibit to teach the public about the Mars exploration rover missions and therefore bears a physical resemblance to the MER (see Figure 1). The PER's camera and infrared rangefinder are mounted on a pan-tilt head, which stands several inches above the main body of the robot. For this study, the PER was run in an obstacle avoidance mode. After



Figure 1. The Personal Exploration Rover.

sensing an obstacle, the PER would stop and turn its head to scan its immediate environment for an unobstructed path. Once it found one, it would turn and move in that direction. If there was no clear path, the PER would stop. During forward motion, the PER's head moved back and forth to simulate searching behavior. The PER's camera provided a continuous video feed to a nearby laptop, which was visible to children. See Nourbaksh et al. (2006) for additional technical details.

### *Selecting examples*

The three examples included here are intended to provide descriptive information about children's explorations with a novel robot. They illustrate how children drew conclusions about the PER's capabilities. The examples exhibit some of the most common strategies employed by children interacting with the PER: initiating following-finding behavior (e.g., whistling, clapping hands, or standing in front of the rock), providing directions (e.g., by pointing to the rock or waving to the robot), talking to the robot, and blocking the robot to trigger obstacle avoidance. These examples are illustrative rather than representative of the data as a whole. In addition, these interactions were brief and partially dictated by the robot's capabilities, so it would be inappropriate to generalize from them to all child-robot interactions.

The example write-ups draw on video data of children interacting with the PER, and information from a parent survey of each child's prior experience with robotic technologies.

## **Interactions with the PER**

### *Elizabeth*

Elizabeth,<sup>1</sup> 6 years 11 months, had frequently been exposed to robotic technologies. There were a number of robotic artifacts in her house, including a Robosapien, a toy dog, and some remote-control toys. The family also learned about robots through museum exhibits and books. While Elizabeth was interacting with the PER, her father reminisced about the time they made the Robosapien pet the robotic dog, causing the dog to spin.

Elizabeth believes some robots and computers can have feelings: "sometimes when I put like a disk in, it could feel, feel the brain of the computer." Her father then explained, the family had an old computer that frequently broke down. When it acted up, Elizabeth's father would tell her it was "not happy." When asked wheth-

er she thought that every computer could feel emotions, she responded, “just the one that we have... maybe a couple of them.”

*Elizabeth begins exploring the PER by lying down in front of it to take a closer look. After about 40 seconds, the experimenter asks her if she could get the PER to go to the rock. Elizabeth remains on the floor watching the PER. The experimenter then asks her what would happen if she stood in front of the PER. She gets up to try, but jumps in front of the PER so quickly that its IR sensor doesn't register her. The PER comes so close to hitting her that she steps away and says, “It's gonna crash me.” Immediately after she steps away, the PER stops. Her father points out that the PER has stopped, and she repeats his observation. As the PER turns its head to seek a clear path, Elizabeth points out that it is moving again. Her father reinforces the goal of getting the PER to the rock. In response, Elizabeth briefly gets down on the floor behind the PER, and then runs to stand in front of it.*

*She successfully blocks the forward motion of the PER, and then follows the PER around in a circle to block its path as it turns. Just before it stops, she starts sidestepping away and says, “see if it follows me.” As luck would have it, the PER starts moving in her general direction. She continues to sidestep towards the rock, repeating that it's following her. She stops at the rock, and waits for the PER there.*

*After a short period of time, she moves back to the PER to coax it to the rock by getting in front of it and taking baby steps while saying “come here”; however, the PER has turned to move in the opposite direction. She gets back down on the floor with the PER and starts to gently touch its fake solar panels.*

*She gets back up to try again. She walks over to the rock, but instead of following her, the PER rolls in the opposite direction. She tells it to “stay,” but it doesn't listen. Then she bends down next to the robot and says, “I'm not sure how it's programmed.”*

*Elizabeth's brother joins the interaction after 3 minutes and 40 seconds.*

Elizabeth's exploration of the PER takes her on a circuitous route. She starts out curious about the robot, and then becomes slightly wary of it when it almost runs into her. With her father's encouragement, she reengages it. Elizabeth does not start out believing the PER can follow her, but she spends some time experimenting. Although the confirmatory feedback she receives from the robot is false (it is not following her), it is enough to temporarily validate her belief that the robot has the capacity to follow her. Elizabeth persists in using this strategy to control the robot until it is clear that it is not working. Elizabeth also tried talking to the PER at two points during her interaction, but the robot did not respond to either of her verbal requests.

Throughout this interaction, Elizabeth repeatedly explores the robot's capabilities and then crafts strategies for interacting with the robot based on her beliefs

about those capabilities. Her strategies for engaging the robot are not unlike those a child might try with a reluctant pet: she walks in front of the robot, calls it, gently touches it, and then checks whether it will follow her. It seems that Elizabeth's knowledge of animal behavior is seeping into this interaction, and when she does not know what else to do, she tries an animal-based strategy.

### *Jake*

Jake, 6 years 2 months, had relatively high exposure to robotic technology at home. His toy collection included toys such as Bionicles and Lego Technic, as well as robot videos and remote-control cars. Jake came to the museum with his parents and his younger sister.

*Jake and his sister stand patiently in the middle of the room waiting for the PER to start up. Once it begins moving, the experimenter asks Jake whether he can get the PER to go to the rock. Jake's first move is to wave towards the rock. He then leans towards the PER with his arms outstretched to corral it. Next to him, his sister is imitating his movement.*

*Jake's corralling motion causes the PER to stop its forward movement, and eventually turn away from the rock. In response, Jake points to the rock using vigorous, whole-arm movements. As the robot continues to turn, these movements give way to more targeted finger pointing near the PER's head. Jake's sister is imitating Jake's motion of pointing towards the rock, and although she gets closer to the PER's head than Jake, her actions do not trigger the robot's sensor. Jake eventually comes around to the side of the PER that is farthest from the rock and once again tries to corral it while it is turning. Soon the PER stops turning and resumes forward motion. Jake then steps out of the range of the camera for approximately four seconds, and the PER stops its forward motion, presumably because its IR sensor was triggered.*

*As the PER begins to turn again, Jake and his sister both use their fingers to point towards the rock. After a second or two, Jake again tries to corral the PER while telling it to "keep going" and then "stop, stop!" Jake jumps to the other side of the PER just as it is finishing its turn and resuming forward motion. However, the PER is angled slightly away from the rock, so Jake jumps in front of it to tell it to "turn, turn" while waving his hands in front of the PER's head. After a few waves he holds his hands steady long enough to trigger the PER's sensor, causing it to stop and turn again. As the PER looks around, Jake backs away and stands with his hands outstretched, as if waiting to see what the PER will do. As the PER begins to turn, Jake steps out of its path, and encourages it to "go, go!" while waving it towards the rock. He then bends down and raises his hands next to the PER's head, as if forming a barrier to stop it from turning too far. The PER stops its turn and heads straight for the rock. Jake*

*stands back and puts a hand on his sister to stop her from moving. The PER arrives at the rock after approximately two minutes.*

Jake had a set of reasonable, if not completely accurate, strategies for working with the PER. His main strategy of corralling seemed successful because of two unintended effects: First, the hand gestures associated with corralling often triggered the PER's IR sensor, causing a change in direction. Also, because Jake was corralling the PER towards the rock, he sometimes stood on the side of the PER farthest from the rock, blocking the undesirable direction. However, there were also times when his corralling movements actually moved the PER away from the rock.

It is not clear from Jake's interactions that he understood exactly how to trigger the PER's sensor. Both he and his sister continued to point and wave the PER towards the rock, even though the PER does not respond to either of these actions. Jake also spoke to the PER occasionally, but it is not clear he believed he could control the PER verbally, because gestures accompanied all his speech.

Jake's strategies worked well enough to guide the PER in the short term. He observed the robot's behavior and continued to do what he thought would control the PER. His overall demeanor, including the fact that his often wild gestures became more precise as his hands got closer to the robot, indicated that Jake may be able to collaborate successfully with robots. However, his inaccurate model of the robot's capabilities could be a real problem for long-term collaboration. Pointing when close to the PER's head is counterproductive because it can trigger the IR sensor at inopportune moments. An accurate model of the PER's sensor capabilities is the best way to ensure repeated successful collaborations.

### *Emma*

Emma, 7 years 11 months, was visiting the museum with her mother. Emma has few robotic resources at home. However, she has visited museum exhibits about robots, and she built robots while attending an invention and electronics summer camp.

*Emma begins her interaction with the PER by leaning down and quickly putting her hand in front of its head. However, her action is so quick that it fails to trigger the PER's IR sensor, and the PER makes no response. She asks the experimenter if the PER can turn. The experimenter responds that it can, and instructs her to hold her hand in front of the PER for a little longer to trigger a response. Emma kneels down and holds her hand in front of the PER's head until it starts to turn. She moves her hand away, and exclaims, "hey, cool!"*

*The experimenter then introduces the goal of getting the robot to the rock on the opposite side of the room. Emma comments that the PER is already heading in that*

direction, and she kneels down next to the robot as it turns. She holds her hand off to the side of the robot's head, perhaps to prevent it from overshooting the rock. She moves her hand away once the robot starts rolling forward and gives a small wave in the direction of the rock. The PER moves forward in the general direction of the rock, but Emma becomes concerned it might hit a nearby chair. She gets down on the floor and puts her hand next to the PER's head and then slightly in front of it, but does not hold her hand there long enough to trigger the IR sensor. Her action has failed to deter the PER, and she exclaims, "stop it!" as the PER nears the chair. The experimenter moves the chair out of the way, which causes the PER to stop, turn and head away from the rock.

Once the robot begins moving away from the rock, Emma returns to it. She kneels down on the floor next to the robot, and this time holds her hand in front of the PER's head until she has blocked its forward progress. Emma stands up as the PER begins to turn. The experimenter comments that she just saw a picture of Emma's hand on the laptop. Her mother repeats the comment by saying, "if you put your hand in front of its little sensor again, you can see your hand on the computer," but Emma is pre-occupied with the robot approaching the rock. The PER arrives at the rock shortly thereafter, approximately two minutes after the interaction started.

The experimenter asks Emma if she would like to send the PER somewhere else, and Emma responds by saying, "let's see if it can go to my mommy." But as the PER approaches her mother, Emma runs over to the robot to block it with her hands, commenting that she didn't want the robot to run into her mother. Her mother responds by saying, "I think that's what that little face is all about, so it doesn't run into things." The PER is now moving towards the chair. Emma's mother reaches over to block the PER with her hand, and then tries to block it again as it turns. Emma gives her mother the instruction to "wait till it stops," and then kneels down next to the robot to block it with her own hand. As the robot's head turns, Emma laughs and says, "It doesn't know where it wants."

The experimenter again comments that the laptop is displaying pictures of objects, such as the rock. Emma briefly walks over to the computer screen, then walks back to the middle of the room and says, "I wonder if it will come towards me," but the rover is turning away from her. She kneels down and puts her hand in front of the robot to block it. As she steps away, the robot begins to move towards her. Emma comments, "hey, it's coming towards me," but then she steps away to let the robot "go straight."

The experimenter engages Emma in conversation by asking if she thought the robot would ever bump into anything. Emma replies, "it probably could have bumped into that chair [with skinny legs]... but not solid things like walls or ceilings. It wouldn't bump into a chair like this [points to thick armchair], it's too solid." When asked why that would make a difference, Emma replied, "because this is solid, it could definitely see it." Emma spends the remainder of her encounter watching the

*robot, blocking it with her hands, and predicting its next destination. The interaction ends after nine minutes.*

Emma's explorations of the PER's capabilities yield useful information. After some instruction, Emma learns that she must leave her hand in front of the PER for a certain amount of time to trigger a response. Following this realization, she spends the remainder of the interaction further exploring the limits of the PER. At numerous points she makes astute and correct observations about the PER, such as her comment that the robot would not bump into "solid things" but could get tripped up by thin chair legs. Her instruction to her mother to wait until the robot stopped turning was also a good observation: The PER does not accept any sensor input while it is turning towards the unobstructed path, but most children fail to notice this feature and continue to block the robot while it is turning.

Unlike Elizabeth, Emma never adopts a biological model of the PER. She does not assume that it can hear or follow her. Instead, she treats the PER as mechanical and spends most of the time exploring and speculating about its capabilities. Emma did not have humanlike robots in her house; her experience with robots was based on building them at camp. However, it is difficult to know exactly how her prior experiences contributed to her interaction style.

Of the three children, Emma was the most successful at collaborating with the PER mainly because she was able to figure out the limits of the PER's autonomy and tailor her behavior to complement the robot's abilities. The information she gathered about the PER came from a number of sources, including her mother, the experimenter, and her own observations. Her ability to use different resources to determine the most effective pattern of engagement could have implications for a possible long-term relationship. If given regular opportunities to interact with the robot, one could imagine a successful collaboration developing.

## Discussion

What are the factors that lead to successful human-robot interactions? As our three examples demonstrate, a human user's beliefs about a robot impact the success of the human-robot system. Elizabeth believed the PER would follow her lead, which led her to initiate several unsuccessful following-finding episodes. Jake was mostly successful in his collaboration with the PER, although his strategies, which were based on an inaccurate model of the robot's capabilities, occasionally worked against him. Emma understood what she needed to do to trigger a response from the PER and altered her behavior accordingly, for example, by holding her hand in front of the sensor for a longer period of time.

Earlier in the paper we asked which user beliefs were most likely to predict collaborative success and suggested that accurate beliefs about a robot's autonomous capabilities were important. The examples support our hypothesis that an accurate understanding would strongly affect interaction success. Jake is noteworthy for his strategies, which were workable in the short-term but based upon incorrect assumptions. We suggest that his strategies may not be reliable in the long-term.

Human beliefs about robots come from a number of different sources. The form of the robot is one important source (Kiesler & Goetz, 2002; Woods, Dautenhahn & Schulz, 2004), but the user's experience with robots (Bernstein & Crowley, under review) and unrelated needs have also been shown to influence ideas about robots (Turkle, Taggart, Kidd & Daste, 2006; Turkle, 2007). Given this combination of influences, we should expect that individuals come to an interaction with a unique set of beliefs about robots, some of which may be resistant to change. Emma's experience at electronics and invention camp may have provided her with knowledge about robots that she was able to apply to the PER. It is possible that certain types of experience are more useful than others in providing generalizable models of robots. This would be an interesting question to follow up. But the larger point is that no user will come to an interaction without beliefs about how they might best negotiate with their robot partner. This is why we believe that the robot has a role to play in helping the user learn about its capabilities.

A well-designed robot that can facilitate accurate beliefs about its capabilities will go a long way towards improving its relationship potential. To further this goal, we propose three design guidelines: diagnostic transparency, predictive transparency, and simplicity.

First, a robot should be designed to increase diagnostic transparency (Nourbakhsh, 2000). This means that a user who is unfamiliar with the technological layout of the robot should be able to infer why the robot is not behaving as desired. For example, it is easy for museum employees to determine when the PER's batteries are running low because it arches its head straight up, which it never does otherwise. Diagnostic transparency allows users to accurately establish behavioral causality simply by observing the robot. If we expect users to engage in productive relationships with robots, we must give them a means of interpreting the robot's behavior. One way to ensure that humans correctly assign causality is to have the robot respond in a time-sensitive manner. Human responses are rapid, so having a robot respond quickly may help the user determine cause and effect relationships.

Related to the notion of diagnostic transparency is that of predictive transparency, or the ability of the user to infer what the robot would do under novel conditions. In other words, does the user have an accurate enough mental model of the robot to predict its behavior? This type of transparency allows users to plan scenarios that maximize the autonomy of the human-robot system. For example,

the Roomba user who arranged the furniture around a particularly dirty area used his knowledge of the Roomba's behavior to enhance the system's output; he knew that if he provided the Roomba with boundaries, it would continue to vacuum the spot he wished to keep clean. Predictive transparency is more likely when the robot's behavior can be easily linked to a causal stimulus (i.e., it has high diagnostic transparency). For example, after briefly playing fetch with AIBO, a user can assume AIBO will try to retrieve the ball next time it is thrown. Although the cause and effect relationship is also supported by a metaphor in this case (people expect a dog to fetch a thrown object), diagnostic transparency will generally aid predictive transparency.

Finally, we note that both diagnostic and predictive transparency are enhanced by keeping robot designs as simple as possible, while maintaining the desired functionality. It takes a user longer to create an accurate model of a complex technology than a simple one. Although there are numerous instances of users mastering complex technologies, the complexity should be necessary to the purpose of the robot. A simple robot is more likely to facilitate transparency than a complex one.

Kahn et al. (2007) suggest the development of psychological benchmarks that focus on the robot's ability to elicit the kinds of responses that people typically make with other people. They assume that implementing humanlike characteristics is a gateway to more effective human-robot interaction, because humans are most comfortable with humanlike agents. We have argued for a new approach that focuses attention on the human-robot interaction rather than the human and robot as individual agents. We have introduced the notion of relationship potential as a way to understand the success of human-robot collaborations. The chance of success is much higher if the relationship itself functions with high autonomy and, if autonomy is to be used as a benchmark, it should be measured system-wide. Our approach suggests building robots that help users form accurate mental models of their capabilities by increasing the transparency of the causes underlying the robot's actions (Stubbs, Bernstein, Crowley & Nourbakhsh, 2005).

Before concluding, we pause to consider how our findings might generalize given their focus on children. Children may be less advanced than adults in how they think about and understand robots. However, we have found little evidence to support this. Research comparing children's and adults' beliefs about robots has sometimes found similarities between the two groups. Van Duuren and Scaife (1996; Scaife & Van Duuren, 1995) found that adults and children over age 7 showed similar patterns of responses when asked about characteristics of robots such as the presence of a brain or their ability to perform brain-related tasks. We also know from educational research that children are capable of thinking and learning about complex robotics concepts, such as autonomy (Nourbakhsh et al.,

2005, 2006). Some researchers have even suggested that children are more flexible in their thinking about technology, because they have grown up with it (Turkle, 1984, 1999, 2007). We would further argue that consideration of children is relevant to a broader benchmark discussion because children are an active robotics user group. If psychological benchmarks are not meant to include a significant segment of the user population, it would be good to be explicit about those boundaries early in the development of the concept.

Will better collaboration come from humanlike robots or robots who like humans? In the same way that dogs have been bred to be working companions who respond to training and provoke emotional responses from their trainers, can robots be built so that their human partners want to invest time, emotion, and energy in learning how to build a working relationship? While we have suggested a number of robot features that we believe will enhance the likelihood of productive and long-lasting human–robot relationships, verifying the importance of these features will require long-term studies on both human–robot interaction and the cognitive changes experienced by human users. Such studies will likely require interdisciplinary collaboration, with teams from the learning sciences and robotics challenging each other to rethink how we benchmark and design for human–robot collaboration.

## Acknowledgments

The authors would like to thank Catherine Eberbach, Sasha Palmquist, and Kristen Stubbs for their valuable feedback on this manuscript. We would also like to thank Emily Hamner for her help with the PER, Anuja Parikh for her help with coding, Lowell Schwartz for his help transporting the PER, and the staff and visitors at the Children's Museum of Pittsburgh.

## Note

1. All names are pseudonyms.

## References

- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences*, 9(4), 403–436.
- Bernstein, D., & Crowley, K. (under review). *Searching for signs of intelligent life: An investigation of young children's beliefs about robot intelligence*.
- DiSalvo, C. (2006). *Discovering products in contemporary robotics: Towards a theory of product in design*. Unpublished doctoral dissertation, Carnegie Mellon University.

- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2002). *A survey of socially interactive robots: Concepts, design, and applications* (Tech. Rep. No. CMU-RI-TR-02-29). Pittsburgh, PA: Carnegie Mellon University.
- Forlizzi, J., & DiSalvo, C. (2006). Service robots in the domestic environment: A study of the Roomba vacuum in the home. *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, (pp. 258–265). Salt Lake City, Utah.
- Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19(3), 265–288.
- Kahn, P. H., Freier, N.G., Friedman, B., Severson, R.L., & Feldman, E.N. (2004). Social and moral relationships with robotic others? *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, (pp. 545–550). Okayama, Japan.
- Kahn, P. H., Ishiguro, H., Friedman, B., & Kanda, T. (2006). What is a human? — Toward psychological benchmarks in the field of human-robot interaction. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*, (pp. 364–371). Hatfield, UK.
- Kahn, P. H., Ishiguro, H., Friedman, B., Kanda, T., Freier, N.G., Severson, R.L., & Miller, J. (2007). What is a human? — Towards psychological benchmarks in the field of human-robot interaction. *Interaction Studies*, 8(3). (This issue)
- Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1–2), 61–84.
- Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 576–577). Minneapolis, Minnesota.
- Litman, D. J., Rose, C.P., Forbes-Riley, K., VanLehn, K., Bhembé, D., & Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education*, 16(2), 145–170.
- MacDorman, K. F., & Cowley, S.J. (2006). Long-term relationships as a benchmark for robot personhood. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 378–383). Hatfield, UK.
- Minato, T., Shimada, M., Ishiguro, H., & Itakura, S. (2004). Development of an android for studying human-robot interaction. *Proceedings of the International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems* (pp. 424–434). Ontario, Canada.
- Nourbakhsh, I.R. (2000). Property mapping: A simple technique for mobile robot programming. *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence (AAAI/IAAI 2000)* (pp. 840–845). Austin, Texas, USA.
- Nourbakhsh, I., Crowley, K., Bhavé, A., Hamner, E., Hsiu, T., Perez-Bergquist, A., Richards, S., & Wilkinson, K. (2005). The robotic autonomy mobile robotics course: Robot design, curriculum design and educational assessment, *Autonomous Robots Journal*, 18(1), 103–127.
- Nourbakhsh, I., Hamner, E., Ayoob, E., Porter, E., Dunlavey, B., Bernstein, D., et al. (2006). The personal exploration rover: Educational assessment of a robotic exhibit for informal learning venues. *International Journal of Engineering Education, Special Issue: Trends in Robotics Education*, 22(4), 777–791.
- Scaife, M., & van Duuren, M. (1995). Do computers have brains? What children believe about intelligent artifacts. *British Journal of Developmental Psychology*, 13(4), 367–377.
- Schunn, C. D., Crowley, K., & Okada, T. (2005). Cognitive science: Interdisciplinarity now and then. In S. J. Derry, C. D. Schunn & M. A. Gernsbacher (Eds.), *Interdisciplinary collaboration: An emerging cognitive science*. (pp. 287–315). Mahwah, NJ: Erlbaum.

- Stubbs, K., Bernstein, D., Crowley, K., & Nourbakhsh, I. (2005). Long-term human-robot interaction: The personal exploration rover and museum docents. *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, (pp. 621-628). Amsterdam, Netherlands.
- Stubbs, K., Bernstein, D., Crowley, K., & Nourbakhsh, I. (2006). Cognitive evaluation of human-robot systems: A method for analyzing cognitive change in human-robot systems. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*, (pp. 59-65). Hatfield, UK.
- Thrun, S. (2004). Toward a framework for human-robot interaction. *Human-Computer Interaction*, 19, 9-24.
- Turkle, S. (1984). *The second self: Computers and the human spirit*. New York: Simon and Schuster.
- Turkle, S. (1999). What are we thinking about when we are thinking about computers? In M. Biagioli (Ed.), *The science studies reader* (pp. 543-552). New York: Routledge.
- Turkle, S., Taggart, W., Kidd, C.D., & Daste, O. (2006). Relational artifacts with children and elders: The complexities of cybercompanionship. *Connection Studies*, 18(4), 347-361.
- Turkle, S. (2007). Authenticity in the age of digital companions. *Interaction Studies*, 8(3). (This issue)
- van Duuren, M., & Scaife, M. (1996). "Because a robot's brain hasn't got a brain, it just controls itself" — Children's attributions of brain related behaviour to intelligent artifacts. *European Journal of Psychology of Education*, 11(4), 365-376.
- VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., et al. (2005). The Andes physics tutoring system: Five years of evaluations. *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, (pp. 678-685). Amsterdam, Netherlands.
- Wada, K., & Shibata, T. (2006). Robot therapy in a care house: Results of case studies. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*, (pp. 581-586). Hatfield, UK.
- Woods, S., Dautenhahn, K., & Schulz, J. (2004). The design space of robots: Investigating children's views. *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, (pp. 47-52). Okayama, Japan.

### Authors' addresses

Debra Bernstein and Kevin Crowley  
Learning Research and Development Center,  
1st Floor  
3939 O'Hara Street  
Pittsburgh, PA 15260 USA  
dlb36@pitt.edu; crowleyk@pitt.edu

Illah R. Nourbakhsh  
Robotics Institute, NSH 3115  
Carnegie Mellon University  
Pittsburgh, Pennsylvania 15213 USA  
illah@cs.cmu.edu

### About the authors

**Debra Bernstein** is currently a doctoral student in Cognitive Psychology and a graduate student researcher at the University of Pittsburgh Center for Learning in Out-of-School Environments. She received her M.A. in Educational Psychology from Columbia University in 2002. Her research interests include technology learning in out-of-school settings and the impact of technology on cognitive development.

**Kevin Crowley** is an Associate Professor of Cognitive Studies and Cognitive Psychology and Director of the University of Pittsburgh Center for Learning in Out-of-School Environments. He received his Ph.D. in Psychology from Carnegie Mellon University in 1994. His research interests include family learning and science and technology learning in out-of-school environments.

**Illah Nourbakhsh** is Associate Professor of Robotics in The Robotics Institute at Carnegie Mellon University, Director of the Center for Innovative Robotics and Director of the Community Robotics, Education and Technology Empowerment Laboratory. He received his Ph.D. in Computer Science from Stanford University in 1996. His current research projects include educational and social robotics and community robotics.