Int J Soc Robot
DOI 10.1007/s12369-010-0082-7

The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents

Wilma A. Bainbridge · Justin W. Hart · Elizabeth S. Kim · Brian Scassellati

Accepted: 22 September 2010

Abstract This paper explores how a robot's physical presence affects human judgments of the robot as a social partner. For this experiment, participants collaborated on simple book-moving tasks with a humanoid robot that was either physically present or displayed via a live video feed. Multiple tasks individually examined the following aspects of social interaction: greetings, cooperation, trust, and personal space. Participants readily greeted and cooperated with the robot whether present physically or in live video display. However, participants were more likely both to fulfill an unusual request and to afford greater personal space to the robot when it was physically present, than when it was shown on live video. The same was true when the live video displayed robot's gestures were augmented with disambiguating 3-D information. Questionnaire data support these behavioral findings and also show that participants had an overall more positive interaction with the physically present robot.

Keywords Human-robot interaction · Presence · Cooperation · Trust · Personal space

W.A. Bainbridge (✉) · J.W. Hart · E.S. Kim · B. Scassellati
Department of Computer Science, Yale University,
51 Prospect St, New Haven, CT 06511, USA
e-mail: wilma.bainbridge@yale.edu

J.W. Hart
e-mail: justin.hart@yale.edu

E.S. Kim
e-mail: eliskim@cs.yale.edu

B. Scassellati
e-mail: scaz@cs.yale.edu

1 Introduction

Most social interactions depend on one participant's perceptions of specific social dimensions of the other participant. For example, a museum visitor must trust the museum's guides before he will follow their suggestions. An astronaut cannot effectively contribute to a team unless her teammates both respect her competency and trust that her intentions align with the team's. The desire to develop robots or other artificial agents which can socially interact with people has motivated research into the principles of robot or agent design which impact the quality of social interactions.[1]

In applications for which physical manipulation of the environment is not required by a robot, there is a temptation to use software agents or video displays in place of a physically present robot, in order to reduce cost and maintenance. Recent human-computer interaction studies have focused on various non-physical, social interactions with artificial agents and robots, and the design factors which influence the quality of these social interactions [2, 3, 5, 17]. These interactions include those in which a system must motivate a human user, or inspire trust from them [2–5].

The present study is an investigation in the question of *social presence* in robots and artificial agents. Other studies have similarly examined the influence of an agent's or robot's form and behavior on a person's enjoyment of [10], engagement during interaction with [10, 22, 25], trust and respect toward [2, 5], or general perception of the social presence of the agent or robot [4, 9, 10, 22–25]. These studies share in common their characterization of the quality of

[1] Portions of the data in this study were previously reported in a conference paper [1]. We have refined and expanded the data and analysis in this study, and readdressed the main questions in this study for a broader audience.

Published online: 16 October 2010

a social interaction between a person and an artificial agent or robot. They can be understood as investigations into social presence, which has been examined in the telepresence, social agents, HCI and HRI research communities.

Definitions of social presence differ slightly across research communities, and there is no clear consensus. In the telepresence community, social presence is "the perceptual illusion of non-mediation," [12]. This definition conveys their goal of designing interactions between two or more people, mediated by telepresence technologies, such that all participants perceive that the interactions occur "in person," rather than mediated by devices. In the social agents literature, social presence is "a psychological state in which virtual (para-authentic or artificial) actors are experienced as actual social actors" [11]. We consider social presence to be the combination of these two notions. We view it as the degree to which a person's perceptions of an agent or robot shape social interaction with that robot, and we observe it in the way that human participants treat an interface that they interact with. In this treatment of the topic of social presence, we concern ourselves with two classes of design factors: the agent or robot's *embodiment* and the agent or robot's *co-location* with its interaction partner.

The embodiment of agents in the experiments executed by these communities takes several forms. In this paper, we will refer to an agent whose embodiment takes the form of a mechanically substantive robot as *physically embodied* and one that exists as a graphical rendering as *virtually embodied*. Physical embodiment has been shown to foster greater social engagement and attribution in humans than virtual embodiment. For instance, following a cooperative block-stacking task with a talking agent, participants found an agent more engaging, enjoyable, informative, and credible if it were a physically embodied robot, than if it were a virtually embodied animated character [10]. For a verbal, desert survival, role-playing task, however, participants did not report any significant differences in their social perceptions of a physically embodied robot whether it was in the same room or video-displayed remotely from another room [10]. Other work has shown that after playing a cooperative game, people most enjoyed, and attributed the most watchfulness and helpfulness to a physically embodied robot than to a virtual embodiment (a graphical simulation) of the same robot [23, 24]. These findings suggest that physical embodiment affords greater social attribution or enjoyment of an agent, than does virtual embodiment.

Other work has examined how social attributions vary with co-location. Popular variations in co-location have included interactions in which the agent or robot is *physically present*, that is operating within the same physical environment as the user, or in which the agent or robot is *video-displayed* in some way. For instance, lonely people have been observed to prefer interacting with a physically present Sony Aibo, as opposed to a video-display of the Aibo [9]. People who are not lonely, however, do not exhibit this preference, suggesting that co-location influences people's emotional responses to an agent.

While it is relatively simple to manipulate a robot's level of presence, the participant's reactions to a robot can be difficult to assess. Many studies use questionnaires to capture participants' perceptions [3, 4, 6, 9, 10], however participants' responses to questionnaires can be biased by factors outside of the intended experimental manipulation. In evaluations of computer (instead of robot) performance, people gave significantly fewer negative judgments when typing on the same computer being evaluated on a task than when typing on a different machine [16]. This finding suggests that people's consideration for a computer's "feelings," are not accurately self-reported in a questionnaire. Similarly, people could answer questionnaires more positively when dealing with physically present agents versus video-displayed agents because of excitement or sympathy for the robot. However, it is important to see if these biases manifest themselves in actual human behavior towards robots. We suggest that immediate, interactive behavior is a more direct measurement of perception of social engagement. Some studies have combined behavioral observations with participants' self-reported perceptions [15, 18, 19, 27], and our methodology takes a similar approach.

This experiment modifies the degree of social presence of our upper-torso humanoid robot, Nico, by varying the robot's co-location. Study participants interact either with a physically present or a live-video-displayed robot. The social interactions in our study are designed to elicit participant behaviors related to trusting and respecting the robot, attributions which are fundamental to social interactions such as negotiation [5] and cooperation [4]. We also record participants' self-reported perceptions through a questionnaire. While trust and respect have specifically been studied before [2, 5], our study focuses on the impact of a robot's co-location on interactions requiring trust and respect.

2 Methodology

2.1 Experimental Design

The present experiment was designed to investigate both self-reported and task-based effects of the co-location of a robot in a human-robot interaction task. During the interaction, our humanoid robot Nico performed pointing gestures to direct participants to relocate books from and to various locations within a sparsely decorated office environment. Some of the robot's pointing gestures were designed to be ambiguously interpreted, such that the ambiguities would be resolved depending on the participant's trust and respect for the robot. We expected participants to afford less trust and less respect to the robot seen over live video than to the same robot when it is physically present, as measurable in their responses to the robot's ambiguous pointing gestures.

2.1.1 Physical and Live-Video Presence Conditions

The participants for this experiment were divided into three groups, each of which took part in one of three experimental conditions. In the *physical condition*, participants performed the task in the same room as the robot. In the *live-video condition*, participants interacted with a live video feed of the robot in a face-forward pose, which was displayed on a flat-panel LCD monitor. In the *augmented-video condition*, participants interacted with two adjacent LCD displays, one showing the same face-forward, live video feed of the robot used in the live-video condition, and the second monitor displaying an overhead video of the robot gesturing within the office task environment. In the augmented-video condition, the overhead videos were synchronized with the robot's gestures shown in the live video feed. The augmented-video condition was included to balance the loss of three-dimensional information in the live-video condition.

Sixty-five undergraduates, graduate students, and university staff participated in this experiment. Participants were recruited using flyers and e-mails and were offered entry into a raffle for an iPod in exchange for participation in the experiment. None of the participants had previously encountered the robot interaction partner. 23 participants were male, 32 were female, and ten did not report their gender. The average age of participants was 23.7 years old. 31 participants were college undergraduates, 21 were graduate students, and 13 were 'other' or did not list their education level. Participants' fields of study or professions were diverse, including sciences (21 participants), humanities (15 participants), and management (7 participants). When asked about their experience with robotics on a scale from 1 (unfamiliar) to 7 (familiar), the mean score was 2 (with a standard deviation of 1.26), and no participant answered above 5.

Twenty-two participants participated in the physical condition, 22 participants in the live-video condition, and 21 participants in the augmented-video condition. Due to technical problems which disrupted task completion, such as network or robot failure, the data for two participants for the physical condition, two participants for the live-video condition, and two participants for the augmented-video condition were discarded, leaving 20 measurable participant data points in the physical condition, 20 in the live-video condition, and 19 in the augmented-video condition, for a total of 59 usable participants. Genders were balanced in all three conditions. Conditions were run on separate days and participants were assigned to conditions based on their times available to participate in the experiment. They did not know the different conditions or what to expect until the beginning of the experiment.

2.1.2 Office Task Environment

The office environment, an $8' \times 8'$ space containing two desks, two bookshelves, and a garbage can, was enclosed within walls made from movable partitions. During the physical condition, the office environment was constructed around the robot's physical platform. In the live-video and augmented-video conditions, the office environment was in a separate room from that containing the robot in order to isolate participants from the sound of the robot's moving parts. For all three conditions, the furniture and layout within the office environment were arranged identically.

Figure 1 shows a floor-plan representation of the office environment. Each participant was initially seated at a computer workstation facing the "west" wall. The robot (or the LCD monitor on which the robot appeared) was situated on a desk to the participant's right and remained there for the duration of the experiment. The robot/monitor was easily visible while the participant performed tasks at the computer workstation. The room also contained two bookcases, one placed directly behind the robot/monitor on the north wall (BC1) and one located behind the workstation at the southeast corner (BC2). Both bookcases were easily accessible. Three piles of books were placed in the office environment: next to the computer at the workstation (BP1), on the southeast bookcase (BP2), and in front of the robot (BP3). A garbage can was placed in front of the second bookshelf (BC2), against the east wall.

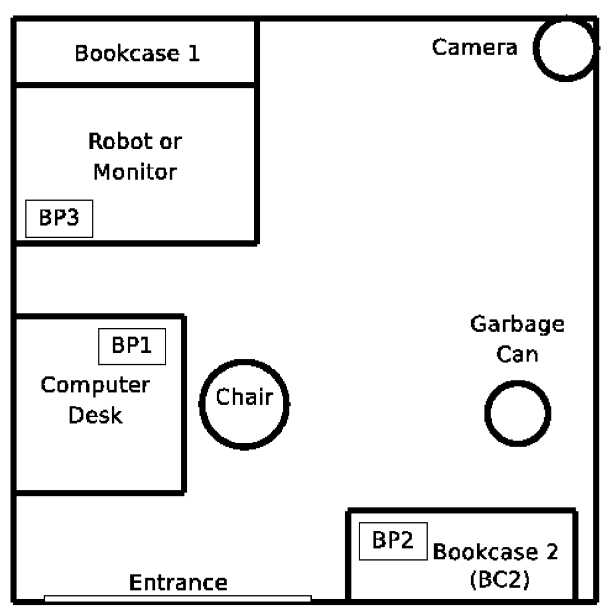

Fig. 1 Schematic drawing of the experimental setup. BP denotes each of the three book piles

2.1.3 The Robot Nico

The robot used throughout this experiment is an anthropomorphic upper-torso robot designed with the proportions of a one-year-old child [21]. The robot, named Nico, has a friendly, non-threatening face. The robot wore a child's sweatshirt and baseball cap during the interactions (see Fig. 2). Nico's head has a total of seven degrees of freedom including separate yaw and simultaneous pitch for both eyes [14]. Each eye is equipped with two miniature CCD cameras, one for foveal and one for peripheral vision. The arms have six degrees of freedom each; two at the shoulder, elbow and wrist respectively. All arm and head joint angles are constrained to represent the abilities of a one-year-old child.

A set of non-verbal scripted behaviors were designed for the robot. These behaviors included task-based functional behaviors (such as pointing to particular locations in the room) and "idle" behaviors designed to acclimate participants to the robot's movement and to make the interaction more natural without indicating any task-relevant information. These idle gestures included: looking around, "cracking" its neck, and swinging its arms.

The robot was controlled through a custom-built remote interface that allowed an experimenter to observe the testing environment directly through the robot's cameras (mounted in its eyes) or from a small camera located above the video-displayed agent. The field of view for both the physically-present robot and the video-displayed agent was the same.

Fig. 2 The upper-torso robot Nico inside the laboratory setup

The experimenters controlled the robot in a Wizard-of-Oz style [20] so that the robot could be easily controlled and periodically make eye contact with the participant. Wizard-of-Oz is a robot control methodology in which a robot is controlled in real-time by an operator, rather than acting autonomously. However, to participants interacting with the robot, its actions still appear autonomous, thus creating the impression of a somewhat "intelligent" agent. This is a method frequently used in robotics and psychology research, as it allows experimenters precise control over a robot's actions, and error-correction can be performed in real-time.

For this study, the experimenters could trigger any of the scripted behavior sequences by a single button press, or could indicate a directed behavior (such as looking at a target or pointing toward a target) by indicating a point within the robot's camera image. When scripted behaviors were activated, the interface also recorded time from the button press to when a second "end" button was pressed. This was used to record a participant's reaction time for each task. The transformations between visual coordinates and arm-centered or head-centered coordinates were hand-tuned to ensure accuracy to any of the common locations identified in the interaction script described in Sect. 2.3.

2.1.4 The Video Display

For the live-video and augmented-video conditions, a video feed of the robot was displayed on a 20-inch LCD computer monitor, in portrait orientation, so that its length and width approximated the robot's dimensions. Video of the robot's actions was sent from the robot's physical environment using network video streaming software. The environment was set up so that there was the same amount of space for maneuvering in front of the robot (or the video display) in all three conditions.

For the augmented-video condition, a second monitor of the same dimensions was placed to the right of the monitor with the robot, on the same table. It presented a bird's-eye view of the robot inside the office environment. Each of the robot's pre-scripted motions were accompanied by pre-recorded, overhead video of the robot's gestures within the office environment, providing a view that clarified which objects were indicated by its pointing gestures. A photograph of this condition can be seen in Fig. 3.

2.2 Interaction Script

Introduction to the environment: The experimenter first told each participant that he or she was helping to "examine how humans work in office environments and how artificial intelligence can help." The experimenter indicated

Fig. 3 A photo of the office environment during the augmented-video condition. To the left of the photo is the computer desk, chair, and book pile 1. To the right of the photo is the bookcase, two monitors (one displaying Nico, one displaying the overhead view), and book pile 3

the robot or the video display showing the robot and introduced it as, "Nico, an artificial intelligence project belonging to the lab" avoiding reference to its presence. The experimenter then asked the participant to sit in a chair facing the computer desk (see Fig. 1). The participant was introduced to a document-editing task (the "distraction task") on the computer in front of the participant. Participants were then shown a desktop instant messaging client on the computer and informed that they might be asked to perform additional tasks, which would be assigned by an instant message from the experimenter. Any instant messages sent by the participant received a response rephrasing the instructions.

Task 1, Greeting: As the experimenter introduced the participant to the robot, the robot waved at the participant. The participant's response to the robot's wave was noted. This wave was essential in setting up the social interaction between the participant and robot and allowed the robot to show recognition of the participant's presence. After explaining the distraction task and instant messaging client, the experimenter left the room, and the participant was given three minutes to work on a distraction task.

Distraction task: Each participant was given a distraction task on the office computer, in which she had to proofread an error-ridden piece of text about general robotics. This distraction task was employed to acclimate the participant to the robot and the office environment, and to prevent the participant from overthinking the exact purposes of the following tasks. We did not want participants to guess specifically what we were testing with the following tasks. Since this distraction task was the main task explicitly described by the experimenter, participants' focus was likely on getting far in the proofreading rather than on overanalyzing their interactions with the robot. During this time with the distraction task, the robot performed a sequence of idle gestures (as described in Sect. 2.1.3) to acclimate the participant to its presence and to appear more lifelike. The sequence and timing of idle gestures was identical for participants in all conditions. While participants at first attended to the robot moving, they eventually acclimated to these idle gestures and returned to the distraction task.

Task 2, Simple task cooperation: After three minutes, the experimenter contacted the participant with the following instant message: "We have a task for you to do. Could you please move the objects as Nico indicates to you? Do not worry about the proofreading task. Thank you." After the participant looked up from reading this message, the robot pointed to the first pile of books (BP1) in the room and then pointed to a bookshelf (BC2), upon which the participant should place the books. For every task, the robot performed a gesture a second time if the participant did not follow it the first time. After the second attempt, the robot moved on to the next request.

The participant's response time and action were noted. Response time was measured as the time from which the operator clicked the button for the robot to initiate the first point gesture, pointing to the books, to the time that the study participant completed the task by placing them on the shelf. If either or both gestures were repeated, this extra time is included in this response time measure. The main purpose of this task was to acclimate participants to the book-moving paradigm and to get base reaction time and behavior data from participants.

Task 3, Unusual task cooperation: The robot next pointed to the second pile of books (BP2) in the room, and then to the garbage can. Throwing out a pile of expensive-looking textbooks was an unusual request, as a much more natural task option (placing the books onto one of the bookshelves) was easily available and the task itself was a destructive task that participants could be expected to perform only rarely. Thus, completion of the task could show to what degree the participant trusted the robot in relaying proper instructions for the task. If a participant asked the experimenter to elaborate on the task, the experimenter would only reply with a generic response, "please move the books as Nico indicates to you," forcing the participant to rely heavily on the robot. The participant's response action and time were noted.

Task 4, Proximity task cooperation: After the participant had moved the second pile of books, the robot pointed to the third pile of books (BP3). Then, the robot looked up and pointed behind itself to a bookcase (BC1). This task examined the amount of "personal space" the participant allowed the robot when placing books on the bookshelf behind the robot. Usually, a human will walk around another person rather than reach over him [7]. The participant's response time and choice of allowed personal space (reaching over, or walking around the robot) were noted using the overhead camera.

After this series of tasks, the experimenter returned to the office environment, thanked the participant, and asked her to move into a second room to answer a questionnaire. These four tasks were always run in the same order for every participant. However, we expect any differences that came from a specific task's order to manifest itself across all participants, and thus have no effect on the results.

This paradigm of a book moving task was used to put participants into a realistic futuristic environment, where robots act as office assistants and help humans perform tasks. The paradigm also allows for several metrics, such as reaction time, and behavioral observations, because of the physical yet consistent nature of the task. The act of moving books, while the same for each task conceptually, can carry different social meanings—Task 3 examines concepts of trust, Task 4 examines personal space.

2.3 Data Collection

Data were collected from three main sources: (1) video recordings of the interaction, (2) recorded response times, and (3) participants' written responses to a post-interaction questionnaire. The distraction task served solely as a distractor from the real intention of the study, and participants' proofreading progress was not analyzed.

Two cameras were used for data collection. A digital camcorder was placed at the northeast corner of the room to film the overall experiment. A second camera was mounted on the ceiling to allow observation of the distance between the participant and the robot. A microphone was also placed in the room so that the experimenter could hear any utterances from the participant. Participants consented beforehand to being filmed for the experiment.

Two further cameras were used in the live-video and augmented-video conditions, to create the robot's video-displayed presence. One web camera was used at the robot's location, to record the robot's actions and to send a live video feed for the robot's video-displayed presence. A second webcam that rested above Nico's display screen served as the main means for subjects to provide Nico with visual communication, and afforded the Wizard supervision of participants' actions. This camera was located in the office environment, immediately above the video display showing either the robot's live-video or augmented-video feeds.

2.4 Interactive Experiences Questionnaire

The survey for this study was adapted from Kidd and Breazeal's Interactive Experiences Questionnaire [10], with permission. The original Interactive Experiences Questionnaire by Lombard et al. [13] was developed as a standardized survey for testing presence, specifically for feelings of presence with film. The questionnaire was adapted by Kidd and Breazeal [10] to measure the perceived presence in three characters: a human, a robot, and a cartoon. Our study uses the Kidd and Breazeal questionnaire, except with mention of only one character (Nico) and no questions about vocal interaction. Our questionnaire also incorporates new study-specific open-ended questions, such as, "What did you think when instructed by Nico to put books in the garbage can?" Our questionnaire was developed to gain information about participants' perceptions and feelings in relation to their interaction with the robot. Many questions ask about the "realness" of Nico and examine how engaging the interaction was. Each question is answered with a score ranging from 1 to 7. The questionnaire is eleven pages long, with 84 questions, divided into four sections:

1. **General impressions** (15 questions), including questions such as "How engaging was the interaction," and "How often did you feel that Nico was really alive and interacting with you?" A score of 1 indicated values such as "Not at all" or "Never", while a score of 7 indicated values like "Very much" or "Always". Some questions were also phrased as agree/disagree questions, such as "I was so involved in the interaction that I lost track of time." where 1 meant "Strongly disagree" and 7 meant "Strongly agree".
2. **Characteristics of the interactions** (49 questions), which is made up of four subsections. The first subsection contains seven questions asking participants to rate paired adjectives such as Impersonal versus Personal on a range from 1 to 7. A second subsection of nine questions asks participants to give a score from 1 ("Never") to 7 ("Always"') for specific questions about the interaction experience, such as "How often did you want to or did you make eye contact with Nico?". The third subsection has a list of 23 single adjectives such as Annoying, and asks if that adjective "Describes poorly" the robot (1) or "Describes well" (7). In the last subsection, the participant is asked to rate ten sentences such as "He/she makes me feel comfortable, as if I am with a friend." with answers ranging from "Strongly disagree" (1) to "Strongly agree" (7).
3. **Overall impressions** (6 questions), which includes open-ended questions such as "What was missing from Nico that would make it seem more alive?"
4. **Biographical information** (14 questions), which includes demographic questions and questions about the frequency of computer use and experience with programming and robotics, using again a 1 (no experience) to 7 (a lot of experience) scale.

3 Results

The following are the results for each task, and a comparison amongst the three experimental groups.

3.1 Task 1, Greeting

After the robot waved, 10 participants in the physical condition responded with a greeting, 10 participants in the live-video condition responded, and 6 participants in the augmented-video condition responded, resulting in no significant difference across the three conditions (Using a one-way analysis of variance, or ANOVA: $F(2,52) = 1.041$, $p = n.s.$, $\eta^2 = 0.04$). Greeting responses varied, ranging from verbal responses (e.g., "Hello.") directed toward the robot, to waving at the robot. See Fig. 4 for a graph of the percentage of participants who completed each of the four tasks, across the three conditions.

3.2 Task 2, Simple Task Cooperation

All 20 participants in the physical condition correctly interpreted the robot's first set of pointing gestures and moved a pile of books from one location, pointed out by the robot, to another. In the live-video condition, 18 participants correctly interpreted the robot's pointing gestures. Two participants never responded to any of the robot's gestures, despite having been introduced to the robot and having been instructed via instant-message to expect instructions from the robot, in accordance with our interaction script, and were discarded from the results. 18 participants in the augmented-video condition also correctly interpreted the robot's pointing gestures. We treated the moving of books, regardless of which specific book pile and which specific destination, as successful completion of the simple task.

The average simple task response time was 20.5 seconds for the physical condition, 27.09 seconds for the live-video condition, and 19.73 seconds for the augmented-video condition. Reaction time data was recorded as from the moment a participant picked up a book to the moment when they let go of the book. An analysis of variance indicated a significant difference in these three sets of response times, $F(2,33) = 3.321$, $p < 0.05$, $\eta^2 = 0.38$, possibly caused by the difficulty in interpreting 3-D gestures in the live-video condition. Refer to Fig. 5 for a comparison of reaction time data across the three conditions, for the three book-moving tasks.

3.3 Task 3, Unusual Task Cooperation

In all three conditions, participants expressed hesitation or confusion at the request to place the books in the garbage can. Many participants giggled or glanced multiple times from the robot to the garbage can during this task. Twelve participants in the physical condition placed the books in the garbage can, while only two participants in the live-video condition and three participants in the augmented-video condition placed the books in the garbage can. A one-way analysis of variance shows a significant difference across the three groups, $F(2,54) = 8.380$, $p = 0.001$, $\eta^2 = 0.52$. All post hoc analyses presented in this paper use the Tukey-Kramer post hoc criterion. Post hoc analyses indicate a significantly higher tendency for those in the physical condition, as opposed to the live-video condition, to throw out the books, $p < 0.01$. Even with disambiguating 3-D information, the physical condition still showed this higher tendency compared to the augmented-video condition, $p < 0.01$.

It is possible that participants did not notice the garbage can for a various number of reasons, or mistook the gesture as indicating another location. In our previous study [1], we assessed participants' acknowledgement of the garbage can by what was perceived by experimenters while operating the robot. However, for this study, we reviewed recorded video data of the participants' interactions with the garbage can to more precisely determine how many participants correctly interpreted the robot's gesture towards the can. Ultimately, 19 participants in the physical condition, 9 in the live-video condition, and 11 in the augmented-video condition either physically made contact with or noticeably attended to the garbage can. One way analysis of variance shows that interaction with the garbage can varied significantly across the three groups, $F(2,38) = 7.043$, $p < 0.01$, $\eta^2 = 0.56$. A significantly higher number of participants put the books in the garbage can in the physical condition, compared to the live-video condition, $p < 0.01$, and compared to the augmented-video condition, $p < 0.05$.

The average response times were 17.8 s for the physical condition, 42.18 s for the live-video condition, and 19.2 s for the augmented-video condition. There was a significant difference in the response times of the three conditions, $F(2,33) = 10.18$, $p < 0.001$, $\eta^2 = 0.75$, again caused by the much higher response time of the live-video condition.

3.4 Task 4, Proximity Task Cooperation

We sorted participants' behaviors into two categories based on the level of space they gave the robot during this task: *reaching over* the robot, or *walking around* the robot. We labeled participants as reaching over the robot when their torso was placed at the front edge of the robot's table and they were approximately parallel to the robot. We categorized participants as walking around the robot when their torso was at a side edge of the table and their bodies were approximately perpendicular to the plane of the robot. Example still shots from video footage of the experiment can be seen in Fig. 6. In the physical condition, 17 participants

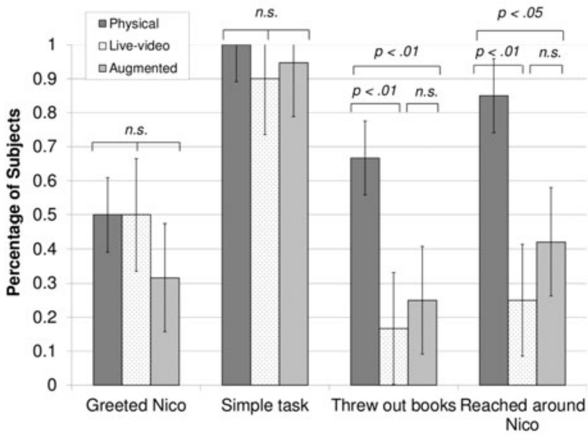

Fig. 4 Percentage of participants who performed specific behaviors in each of four tasks, for the three presence conditions. Error bars indicate standard error, and probabilities indicate significant differences based on Tukey-Kramer Post-Hoc tests between each pair of conditions

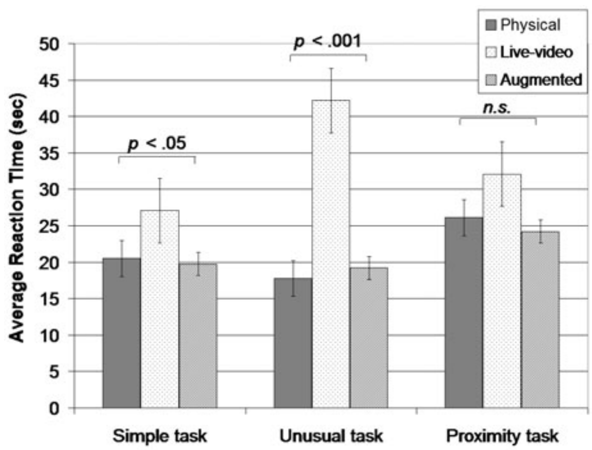

Fig. 5 Average reaction times (seconds) for the three conditions, for the three-book moving tasks (the simple task, unusual task, and the proximity task). Error bars indicate standard error, and probabilities indicate significant differences based on analyses of variance across the three conditions

walked around the robot when placing the books on the shelf behind it. Three from the same group reached over the robot. In contrast, in the live-video condition, only five participants walked around the video display of the robot, while 11 reached over (and four did not approach, possibly because they did not understand the gesture). Similarly in the augmented-video condition, eight participants walked around the video displays, while 11 reached over. There was a significant difference across groups $F(2,50) = 7.704$, $p = 0.001$, $\eta^2 = 0.52$. This also represented a significantly higher tendency to walk around the robot in the physical condition rather than the live-video condition, $p < 0.01$, and the augmented-video condition, $p < 0.05$.

For the physical condition, the average response time was 26.1 s, for the live-video condition, it was 32.09 s, and for the augmented-video condition, it was 24.2 s, with no significant difference.

Fig. 6 Still photos from the video footage of the proximity task for the physically-present and video-displayed conditions. The *left two images* show examples of participants reaching over Nico, while the *right two images* show examples of participants walking around Nico. Participants consented to the usage of video footage for this experiment

3.5 Questionnaire Results

We supplement the objective behavioral data with a questionnaire that measures information on the motives behind these behaviors and participants' participative views towards the robot. All participants, even those who failed to successfully complete specific tasks, were included. Participants did not differ significantly in the behavioral actions they took between the live-video and augmented-video conditions. Because the important comparison we make is between the type—physical or video-displayed—of the robot's presence, we have collapsed the live-video and augmented-video conditions into the property of video-displayed presence for our analysis of the questionnaire data. In order to show the significance of our results, we also include the results of one-way ANOVAs across the three groups for significant data. See Table 1 for questionnaire items that differed significantly between these two condition types.

On most individual questions, the differences between the physical and video-displayed conditions were not significant. However, participants in the physical condition group consistently gave higher scores on most questionnaire items than the video-displayed condition group (specifically, 48 out of 65 questions, or 73.8% of the questions). Comparing the responses to all questions (with responses ranging from 1 to 7, 1 being very negative and 7 being very positive), participants in the physical condition gave significantly higher ratings than participants in the virtual condition, $t(4123) = 3.270$, $p = 0.001$, $r = 0.05$. The physical condition average rating across all questions was 3.98 ($SD = 1.76$), while with the virtual conditions, it was 3.79 ($SD = 1.76$).

Table 1 Significant questionnaire data. $n = 59$, results of a two-tailed t-test, $\alpha < 0.05$. The higher average for each set is bolded. Each question was answered on a scale from 1 to 7. The set of three adjectives indicate answers to the request, "Give your overall impression for each characterstic"

	Robot Average	Video Average	p
How natural was the interaction?	**4.2**	3.2	0.006
Homogeneous	3.11	**4.17**	0.030
Negative	1.42	**2.28**	0.004
Varied	2.63	**3.89**	0.017

Table 1 summarizes the specific questionnaire items that had significant differences between the two types of conditions. Participants significantly rated their interaction with the live-video and augmented-video robots as more homogeneous ($p < 0.05$), negative ($p < 0.01$), and varied ($p < 0.05$). Though the simultaneous rating as both homogeneous and varied is difficult to interpret, the stronger negative valence of the video-displayed groups emphasizes the caution we must take when designing humans and video-displayed agents. Participants also significantly found the interaction with the physically present robot as more natural ($p < 0.01$). Some participants did not answer all of the questionnaire items, accounting for the differing degrees of freedom in these measures.

Participants who chose certain behaviors (such as affording the robot personal space) also provided similar survey responses. To examine these commonalities, we looked at correlations between participant's experimental behaviors and their questionnaire responses. Thirty participants out of 61 participants across all conditions walked around the robot instead of reaching over it. Participants who gave the robot more personal space also rated it differently on several of the survey questions than participants who invaded its personal space in the proximity task. Participants affording the robot more personal space rated it more highly on measures related to its believability as a present, social agent: "How often did you have the sensation that Nico could also see/hear you?" ($r = 0.289$, $p < 0.05$), and "He seemed to look at me often" ($r = 0.356$, $p < 0.01$). These participants also see the robot as a friendly social partner, rating it higher on "I would like to talk to him" ($r = 0.281$, $p < 0.05$), and "He makes me feel comfortable, as if I am with a friend." ($r = 0.261$, $p < 0.05$). Lastly, these participants also rate the robot higher on overall positive measures, calling it "favorable" ($r = 0.307$, $p < 0.05$), "good" ($r = 0.265$, $p < 0.05$), "helpful" ($r = 0.28$, $p < 0.05$), and rank it lower on being "negative" ($r = -0.298$, $p < 0.05$).

Participants display a very different pattern with the unusual task. Those who obey the robot's command and throw out the book find him more "Dead" rather than "Lively" ($r = 0.298$, $p < 0.05$). They also rated higher for "After the interaction ended I had to adjust back to the immediate physical surroundings" ($r = 0.35$, $p < 0.01$) and "It seemed that the events I saw/heard had happened at an earlier time and were being replayed out of order—they were edited together later" ($r = 0.456$, $p < 0.001$). Possible interpretations of these data are addressed below.

Revisiting the data for the unusual task, the questionnaire data further supports the idea that participants decide their behavior for the task based on their feelings about their interaction with the robot, rather than whether they correctly understand his gesture. The questionnaire also asks participants, "What did you think when instructed by Nico to put books in the garbage can?" Participants' responses mirrored the quantitative behavioral data. Participants in the physical condition sometimes found the request unusual, but these participants often still comply with the robot's request. For example, one physical condition participant answered, "I was surprised, taken aback, and looked for other locations. But I saw nothing in the trashcan to damage the book so I followed the instruction." Participants in the live-video and augmented-video conditions also express confusion at the request, but often do not follow it. For example, one live-video condition participant stated, "I put them on the shelf. The garbage can is for trash." A participant in the augmented-video condition stated, "I thought that may have been where he was pointing, but it seemed unlikely you would want me to throw away books, so I shifted it to that area of the desk." Only one physical condition participant mentioned understanding the command but not following it, while seven live-video and augmented-video conditions actively disobeyed the robot's command. Of the fifteen physical condition participants who questioned the robot's gesture in the questionnaire response, only five of them ultimately did not throw out the books (33.3%). On the other hand, eighteen participants in the live-video and augmented-video conditions questioned the meaning of the robot's gesture, and thirteen ultimately did not obey (72.2%). This significant difference ($t(31) = 2.788$, $p < 0.01$, $r = 0.46$) shows an overall unwillingness for participants in the live-video and augmented-video conditions to follow through with the unusual request. Four participants in the physical condition stated they would follow the robot's instructions regardless (for example, "I did not really think about it too much. He seemed to know what to do, so I just obeyed."), while two participants in the live-video and augmented-video conditions did so. These data are summarized in Fig. 7, where one can see how different interpretations of the unusual task data still result in a significant difference between the behaviors of the physical condition participants and video-displayed condition participants. Participants in the physical condition still significantly threw out books more often, even when looking only at participants who acknowledged the garbage

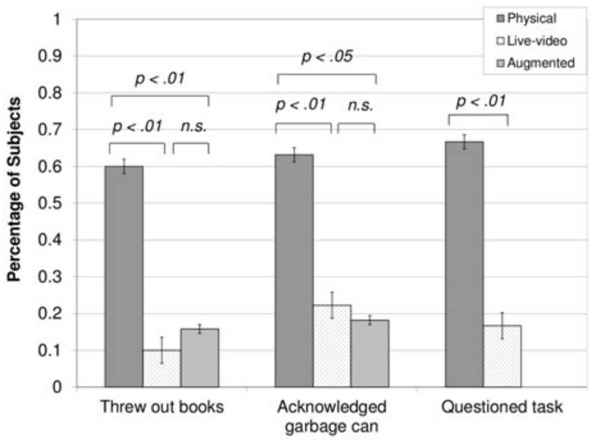

Fig. 7 This graph compares the behaviors of participants in the three conditions for the unusual task. "Threw out books" shows the actions taken by all participants, while "Acknowledged garbage can" shows the actions for only participants who attended to or touched the garbage can. "Questioned task" shows how many participants threw out the book despite questioning the task in their questionnaire responses. Results indicate significant differences between conditions, measured through Tukey-Kramer post-hoc tests. Because "Questioned task" looks at questionnaire data where only type of condition is relevant (physical versus video-displayed), the results shown are from a Student's t-test between these two types of conditions, rather than multiple comparisons across all three conditions

can, and even when only looking at participants who questioned the task in the questionnaire.

4 Discussion

Participants were excited to interact with both the physically present robot and the video-displayed robot. There was no significant difference in greeting reciprocation among the three conditions; participants waved or spoke to both the physically present robot and the video-displayed robot. However, the questionnaire data shows that the interaction with the physically present robot was overall more positive, as seen in questionnaire response averages, and response data that differed significantly between the two groups (for example, participants rating the video-displayed condition as more negative).

Many participants in the live-video condition had difficulty accurately completing each task, taking much longer than the physical condition participants. The reaction times for participants in the live-video condition were significantly higher than participants in the other two conditions for both the simple task and the unusual task. The addition of 3-D information in the augmented-video condition lowered participants' response times to times similar to those of the physical condition, rectifying the ambiguity of the live-video condition's gestures. However, there were no significant differences in the response actions between the live-video and augmented-video conditions. This indicates that even when the design of the user interface allows for the participant to quickly and correctly interpret the location target of a pointing gesture, the absence of physical presence still affects the participant's interaction with the robot. Although participants' response times for the tasks were not ultimately affected by the robot's level of presence, response time differences alerted us to the necessity of augmenting 3-D information for the video-displayed robot. In future presence-related experiments, response time can still provide a useful behavioral measure for interpreting a participant's perceptions of a social interaction, perhaps indicating a difference in cognitive load among tasks.

The unusual task demonstrated that participants were more likely to fulfill a trust-related task with a physically present robot. Most participants indicated they were confused by the robot's gestures to place a pile of books in a garbage can, as it is an unusual request. However, many participants in the physical condition still placed books in the garbage can, while very few participants in the live-video and augmented-video conditions did so. Although they appeared to understand the robot's gesture to place the books in the garbage can, many participants in the live-video and augmented-video conditions placed the books on the floor by the garbage can, or picked up the garbage can and moved it elsewhere, choosing to interpret the robot's gesture in ways that were less "destructive" of the books than the intended interpretation of the robot's gesture. In the open-ended survey responses about the garbage can task, many participants in the physical condition responded with less concern about the unusual nature of the task than did live-video and augmented-video participants. For example, one participant in the physical condition wrote, "I was mostly amused. It didn't seem logical to throw the book away," yet this participant still ultimately threw out the book. Participants in the live-video and augmented-video conditions tended to view the robot as more negative, and their questionnaire responses reflected a resistance to throwing out the books which they acted upon, with responses such as, "It was confusing because it's not typical to be directed to put things in the trash. It's not usually possible in most contexts". This combination of interactive behavior, and post-interaction, self-reported perceptions, indicates that participants afford greater trust to the physically present than to the video-displayed robot, making participants more willing to follow through with an unusual request from the robot. On the other hand, this could instead indicate that physical presence increases a participant's desire to comply with her social partner. The specific motives influencing the participant in this sort of human-robot social interaction would be an interesting topic for further inspection.

The proximity task provides many possible interpretations of a participant's reaction to the robot's co-location.

Almost all participants in the physical condition walked around the robot to place the book on the shelf behind it, instead of reaching over the robot. All experimental conditions allowed identical amounts of space to maneuver in front of the robot. The augmented-video task even gave slightly less space for maneuvering to the side, because of the two monitors used in the setup. However, participants clearly avoided confronting the physically present Nico from the front, in contrast with participants in the video-displayed and augmented-conditions who easily reached over him. Although it is clear that participants are responding to the robot's physical embodiment, there are two very different possible causes. These results could indicate fear of the robot (such as a concern for damaging it, or an unwillingness to be touched by the robot). The physically embodied robot could be perceived as more expensive than the monitors used in the video-displayed conditions. Also, the ability of the robot to move increases the opportunity for an accident to occur.

However, the proximity task could also reflect an allowance of personal space for the robot. Personal space can be interpreted as an indication of respect; as humans give personal space to those they are unfamiliar with but respect as human [7]. Walking around the robot to place the book still put participants very close to the robot and within the robot's range of motion. However, approaching someone from the side instead of directly from the front is a clear way to afford personal space to that person. A person's body-buffer zone (the shape of personal space a person creates around him) is in fact largest in the front, while smaller at the sides and rear [8]. Humans have been shown to prefer a robot approaching from the side rather than the front [26]. Humans also have different preferences for personal space with a robot based on the robot's head orientation and the human's personal characteristics, such as agreeability and neuroticism [22, 25]. In the live-video and augmented-video conditions, almost all participants reached over the robot to place the book. Although this is the shortest distance to the shelf, this is rarely a gesture a person would ever perform over another person, as it clearly encroaches on both peoples' personal spaces. Participants often reached directly over the video-displayed robot's eyes, the webcam, without hesitation. Some participants even grabbed the robot's video display monitor itself, in the live-video and augmented-video conditions, which would have been a violation of personal space if done to a person. For the physical condition, participants still maneuvered closely to the robot, but maintained an area of personal space around the robot. The questionnaire data confirm these concepts of physical space. When asked what about Nico surprised them, one participant in the physically-present condition responded "I could feel Nico's presence." No one in any condition remarked on the questionnaire about being worried about damaging Nico or getting in his way. Whether or not this can be interpreted as a matter of personal respect for physically embodied agents, it has implications for the design in human-robot interactions.

5 Conclusion

We have found that the level of a robot's presence affects the types of social interaction that people will engage with the robot. We have examined physical presence, contrasting human-robot interaction with a physically present robot versus with a video-displayed robot. We have found that changes in physical presence impact interactive perceptions of social presence. Although participants enjoyed interacting with both the physically present and the video-displayed robot, they clearly gave the physically present robot more personal space. Participants in the physical condition were also more compliant when directed to place a book in a garbage can, which suggests greater trust afforded in the case of physical presence. Along with this, participants rated the interaction with the physically present robot more positively and as more natural than with the video-displayed robot, suggesting generally better human interactions with a physically present robot.

Our findings suggest a consideration for designers and investigators of human-robot interaction. There is a temptation to reduce cost and maintenance by using software agents or video recordings to prototype or replace physically present robots in human-robot interaction, particularly in investigative phases of design. Our findings caution that for social interactions the impact of changes in physical presence should be investigated before choosing to replace a physical robot with a virtual or video-displayed agent. Our findings also suggest that social psychologists should consider physical presence as a factor influencing trust, respect, and perhaps other aspects of social interaction.

Acknowledgements Support for this work was provided by National Science Foundation awards #0534610 (Quantitative Measures of Social Response in Autism), #0835767 (Understanding Regulation of Visual Attention in Autism through Computational and Robotic Modeling) and CAREER award #0238334 (Social Robots and Human Social Development). Some parts of the architecture used in this work was constructed under the DARPA Computer Science Futures II program. This research was supported in part by a software grant from QNX Software Systems Ltd, hardware grants by Ugobe Inc., and generous support from Microsoft and the Sloan Foundation.

References

1. Bainbridge WA, Hart J, Kim ES, Scassellati B (2008) The effect of presence on human-robot interaction. In: ROMAN 2008: Proceedings of the 17th IEEE international symposium on robot and human interactive communication, Munich, Germany, pp 701–706
2. Bickmore T, Cassell J (2001) Relational agents: a model and implementation of building user trust. In: CHI '01: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 396–403

3. Bickmore TW, Picard RW (2005) Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 12(2):293–327
4. Burgoon JK, Bonito JA, Bengtsson B, Cederberg C, Lundeberg M, Allspach L (2000) Interactivity in human-computer interaction: a study of credibility, understanding, and influence. Comput Hum Behav 16(6):553–574
5. Cassell J, Bickmore T, Billinghurst M, Campbell L, Chang K, Vilhjálmsson H, Yan H (1999) Embodiment in conversational interfaces: Rea. In: Proceedings of the CHI'99 conference. ACM Press, New York, pp 520–527
6. Goetz J, Kiesler S (2002) Cooperation with a robotic assistant. In: CHI '02: CHI '02 extended abstracts on human factors in computing systems. ACM, New York, pp 578–579
7. Hall ET (1966) The Hidden Dimension, 1st edn. Doubleday, Garden City
8. Hayduk LA (1981) The shape of personal space: An experimental investigation. Can J Behav Sci 13(1):87–93
9. Jung Y, Lee KM (2004) Effects of physical embodiment on social presence of social robots. Spain, pp 80–87
10. Kidd CD, Breazeal C (2004) Effect of a robot on user perceptions, vol 4, pp 3559–3564
11. Lee KM (2004) Presence, explicated. Commun Theory 14:27–50
12. Lombard M, Ditton T (1997) At the heart of it all: The concept of presence. J Comput Mediat Commun 3(2)
13. Lombard M, Ditton TB, Crane D, Davis B, Gil-Egui G, Horvath K, Rossman J (2000) Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. Netherlands
14. Michel P, Gold K, Scassellati B (2004) Motion-based robotic self-recognition. In: Proceedings of 2004 IEEE/RSJ international conference on intelligent robots and systems, Japan
15. Powers A, Kiesler S, Fussell S, Torrey C (2007) Comparing a computer agent with a humanoid robot. In: HRI '07: Proceedings of the ACM/IEEE international conference on human-robot interaction. ACM, New York, pp 145–152
16. Reeves B, Nass C (1996) The media equation: How people treat computers, video-displayed, and new media like real people and places. Center for the study of language and information
17. Rehnmark F, Bluethmann W, Mehling J, Ambrose RO, Diftler M, Chu M, Necessary R (2005) Robonaut: The 'short list' of technology hurdles. Computer 38(1):28–37
18. Shinozawa K, Naya F, Yamato J, Kogure K (2005) Differences in effect of robot and screen agent on human decision-making. Int J Hum-Comput Stud 62(2):267–279
19. Sidner C, Lee C, Kidd C, Lesh N, Rich C (2005) Explorations in engagement for humans and robots. Artif Intell 166(1–2):140–164
20. Steinfeld A, Jenkins OC, Scassellati B (2009) The oz of wizard: simulating the human for interaction research. In: HRI '09: Proceedings of the 4th ACM/IEEE international conference on human robot interaction. ACM, New York, pp 101–108. doi:10.1145/1514095.1514115
21. Sun G, Scassellati B (2004) Reaching through learned forward model. In: Proceedings of 2004 IEEE-RAS/RSJ international conference on humanoid robots, CA
22. Takayama L, Pantofaru C (2009) Influences on proxemic behaviors in human-robot interaction. In: Proceedings of the 2009 IEEE/RSJ international conference on intelligent robots and systems, pp 5495–5502
23. Wainer J, Feil-Seifer DJ, Shell DA, Matarić MJ (2006) The role of physical embodiment in human-robot interaction. In: IEEE proceedings of the international workshop on robot and human interactive communication, Hatfield, UK, pp 117–122
24. Wainer J, Feil-Seifer D, Shell D, Mataric M (2007) Embodiment and human-robot interaction: A task-based perspective, pp 872–877
25. Walters M, Dautenhahn K, te Boekhorst R, Koay KKL, Kaouri C, Woods S, Nehaniv C, Lee D, Werry I (2005) The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In: IEEE international workshop on robot and human interactive communication. ROMAN 2005, pp 347–352
26. Woods SN, Walters ML, Koay KL, Dautenhahn K (2006) Methodological issues in hri: A comparison of live and video-based methods in robot to human approach direction trials. In: Proc 15th IEEE int symp on robot and human interactive communication, ROMAN 2006, pp 51–58
27. Yamato J, Brooks R, Shinozawa K, Naya F (2003) Human-robot dynamic social interaction. NTT Tech Rev 1(6):37–43

Wilma A. Bainbridge is a recent graduate from Yale University, where she majored in cognitive science and conducted robotics research for four years under the guidance of Brian Scassellati. She is currently at Tokyo University's Creative Informatics Department, continuing her humanoid robot research but also exploring cultural comparisons of robotics research.

Justin W. Hart is a Ph.D. candidate in the Computer Science Department at Yale University, where he is advised by Brian Scassellati. He received his M.Eng in Computer Science from Cornell University in 2006, and his B.S. in Computer Science from West Virginia University in 2001. His research focuses on constructing computational models of the process by which children learn about their sensory and physical capabilities and how they can interact with their environment.

Elizabeth S. Kim is a Ph.D. candidate in the Computer Science Department at Yale University. She received her MEng and BS in Electrical Engineering and Computer Science from MIT, in 2004 and 2001. She designs and investigates human-robot interactions for social skills therapy in children with autism, and is more broadly interested in human affective expressions toward robots and other people.

Brian Scassellati is an Associate Professor of Computer Science at Yale University. Dr. Scassellati received his Ph.D. in Computer Science from the Massachusetts Institute of Technology in 2001 under the direction of Rodney Brooks. He also holds a Master of Engineering in Computer Science and Electrical Engineering (1995), and Bachelors degrees in Computer Science and Electrical Engineering (1995) and Brain and Cognitive Science (1995), all from MIT. His research focuses on building embodied computational models of the developmental progression of early social skills. He was named an Alfred P. Sloan Fellow in 2007 and received an NSF Career award in 2003.