Mental Models of a Mobile Shoe Rack: Exploratory Findings from a Long-term In-the-Wild Study

MATTHEW RUEBEN, University of Southern California, USA
JEFFREY KLOW, MADELYN DUER, ERIC ZIMMERMAN, JENNIFER PIACENTINI, MADISON BROWNING, FRANK J. BERNIERI, CINDY M. GRIMM, and WILLIAM D. SMART, Oregon State University, USA

Most people do not have direct access to knowledge about the inner workings of robots. Instead, they must develop mental models of the robot, a process that is not well understood. This article presents findings from a long-term, in-the-wild, qualitative, hypothesis-generating study of the mental model formation process. The focus was on how (qualitatively) users form mental models of the robot—specifically its perceptual capabilities, rules of behavior, and communication with other humans. Participants of diverse ages had multiple interactions with the robot over six weeks in a non-laboratory setting. The robot's rules of behavior were changed every two weeks. A novel, non-anthropomorphic robot was created for the study with a realistic use case: storing people's shoes during a yoga class.

This article reports findings from a case study analysis of 28 interviews conducted over six weeks with six participants. These findings are organized into six topics: (1) variability in the rate at which mental models are updated to be more predictive, (2) types of reasoning and hypothesizing about the robot, (3) borrowing from existing mental models and use of imagination, (4) attributing sensing capabilities where there are no visible sensors, (5) judgments about whether the robot is autonomous or teleoperated, and (6) experimenting with the robot. Specific suggestions for future research are given throughout, culminating in a set of study design recommendations. This work demonstrates the fruitfulness of long-term, in-the-wild studies of human-robot interaction, of which mental model formation is a foundational aspect.

CCS Concepts: • **Human-centered computing** → *Empirical studies in HCI*; *Empirical studies in interaction design*; • **Security and privacy** → Usability in security and privacy;

Additional Key Words and Phrases: Mental models, long-term studies, natural settings, in-the-wild studies, privacy, transparency, non-anthropomorphic robots, robotic furniture

M. Rueben is located at the University of Southern California, Los Angeles, CA, 90089, USA. He is with the Interaction Lab in the Department of Computer Science.

All other authors are located at Oregon State University, Corvallis, OR, 97330, USA. Klow, Grimm, and Smart are with the Collaborative Robotics and Intelligent Systems (CoRIS) Institute. Bernieri is with the School of Psychological Science. Duer, Zimmerman, Piacentini, and Browning are undergraduate students.

M. Rueben's work was partially supported by an NSF NRI grant for "Socially Aware, Expressive, and Personalized Mobile Remote Presence: Co-Robots as Gateways to Access to K-12 In-School Education," Grant No. NSF IIS-1528121.

Authors' addresses: M. Rueben, University of Southern California, Ronald Tutor Hall, Room 423, 3710 McClintock Ave, Los Angeles, CA 90089; email: mrueben@usc.edu; J. Klow, M. Duer, E. Zimmerman, J. Piacentini, M. Browning, F. J. Bernieri, C. M. Grimm, and W. D. Smart, Oregon State University, 101 Covell Hall, Corvallis, OR 97331; emails: {klowj, duerma, zimmerer, piacentj, brownim2, frank.bernieri, cindy.grimm, smartw}@oregonstate.edu.

2573-9522/2021/02-ART16
https://doi.org/10.1145/3442620

ACM Reference format:
Matthew Rueben, Jeffrey Klow, Madelyn Duer, Eric Zimmerman, Jennifer Piacentini, Madison Browning, Frank J. Bernieri, Cindy M. Grimm, and William D. Smart. 2021. Mental Models of a Mobile Shoe Rack: Exploratory Findings from a Long-term In-the-Wild Study. *ACM Trans. Hum.-Robot Interact.* 10, 2, Article 16 (February 2021), 36 pages.
https://doi.org/10.1145/3442620

1 INTRODUCTION AND MOTIVATION

As complex systems of hardware and software, robots are quite difficult to understand. The people who design, build, and program a robot tend to understand it, because they have insider knowledge about its intended functions and inner workings. They are generally quite aware of what it can sense, how it thinks, the rules governing its behavior, its learning capabilities, and its connectedness to other computer systems. Unfortunately, it is difficult to pass this understanding on to users. Instead, users typically rely on what Don Norman calls the "system image" [34]: the parts of the system that are observable by users, including how it looks and behaves. Users then use these observations to develop a "mental model" of the robot that is used to explain and predict the robot's behaviors [34]. This mental model can include the user's perception of what the robot can sense [24], how it processes information (including memory and learning) [8], its behavior and "personality" [52], its role in its setting [31], what it knows [25], and other abilities and attributes [3].

Understanding how people form mental models of robots is critical for a number of applications. When working with robotic teammates, for example, human partners will need to understand what the robots know and intend to do next [19, 44]. Designers will need to know how the robot can best communicate its sensing capabilities to users—e.g., so they can make better-informed judgments about their personal privacy [42]. It will also be useful to influence users' mental models to encourage them to like the robot and accept it for long-term use [5].

It is not fully understood how people form mental models of robots. Theories of mental models from cognitive science have mostly been applied to how people understand engineered systems [32] and computer systems in particular [37]. Very few published human-robot interaction (HRI) studies have looked at mental model formation over time, and those that do (e.g., Stubbs et al. [49]) seem to focus on quantifying the changes in the mental models instead of describing the process that governs those changes. This study focuses on the process instead. The analysis will show *how* people use their pre-existing mental models as well as observations from multiple interactions with a robot to develop a mental model for it over time.

This article reports findings from six weeks of interactions with our robot, the "mobile shoe rack." The study took place "in the wild" [41]—specifically, in the hallway outside a yoga classroom on a university campus (see Figure 1). Data were analyzed from 28 interviews with six participants. An overview of these data is presented in Section 6. A qualitative, case study approach was used to study how participants updated their mental models of the robot by reasoning from their observations and assumptions. Three aspects of their mental models were studied: what the robot can sense and infer, its rules of behavior, and how connected it is to humans (specifically, whether it was being remote controlled by members of the study team). Relevant passages of the interviews were identified, aggregated, and then compared across interviewees via a cross-case analysis. Findings were then organized into six topics (Section 7), including how people reason and hypothesize about the robot's behaviors, borrow from existing mental models and use their imagination, and judge whether the robot is autonomous or teleoperated.

This article contributes early documentation of phenomena that we recommend as high priorities for HRI research. The second contribution is new research questions about mental model

Fig. 1. Reenactment of a typical hallway scene during the study.

development throughout Section 7 as well as in Sections 8 and 9. Finally, Section 10 contributes recommendations for the design of future studies of mental model formation in long-term human-robot interactions. The main goal of this work is to prepare the way for future research in this relatively new domain.

2 BACKGROUND

2.1 Novelty and Habituation Effects

Characteristics of the early stages of interaction with a robot do not always generalize to later interactions when people have gotten used to it. An early forewarning of this problem was documented with studies of "video telephones" at Bellcore in the early 1990s wherein usage dropped off after a few days [10]. A clear example on a robot is documented for the case of "Valerie the roboceptionist." Though people were still using "Valerie" 9 months after deployment, interactions were longer in the first week [14]. The authors called this a "novelty effect"—i.e., the fact that the robot was *new* caused a temporary increase in people's interest levels [14]. Lemaignan et al. [27] have proposed a theoretical model that explains the novelty effect on anthropomorphism by how people borrow and adapt their existing mental models to understand the robot. A related phenomenon is the "habituation effect," in which the effect of some aspect of the robot on the interaction changes as people get used to the robot. It was observed in one study that participants initially allowed a mechanical-seeming robot to approach more closely than a humanoid-seeming robot, but this effect faded with subsequent interactions [21]. Both novelty and habituation effects show how important it can be to run a longer study to see how the interaction changes over time.

2.2 Mental Models

"Humans construct internal representations, or models, of objects in the environment, such as other people, animals, and machines..." [38]. These mental models are "the conceptual frameworks that support people's predictions and coordination in a dynamic world" [20]. Mental models are often incomplete and inaccurate, and prior research indicates that people hold only primitive mental models of objects, technologies, and ideas that they are not familiar with [13, 38]. People then change their mental model of something based on their interactions with it to make the model more useful for achieving their goals [35]. A mental model is different from a "shared mental model," which is one agent's model of its team and of the team's task [44].

When people create a mental model of a complex system, such as a computerized control interface, those with more accurate mental models are more effective at using the system [18]. Metaphors are often effective for bootstrapping accurate mental models for novice users of systems [1, 2], though improper metaphors may result in inaccurate models that are difficult to adapt, hindering the user's ability to use the system [16]. In some cases a reasonable metaphor for an initial mental model may not come to mind when encountering a novel system, leading to difficulty interacting with the system or describing its workings [22].

Several ways to gather information about someone's mental model of a robot have been documented. One strategy is to ask direct questions about the robot and its behaviors: e.g., "Describe the robot's task"; "Why does the robot stop?"; and, "Why do its lights flash?" [52] The language someone uses to describe the robot and its actions can also be informative. For example, mentioning that the robot is aggressive implies it has personality traits and saying that the robot is angry implies it has emotional states, whereas saying "the robot hit me" implies neither [12]. These former two responses also suggest the person is using the intentional stance to understand the robot [6] as opposed to the physical stance or the design stance [7]. It may also be possible to measure a person's experience of "cognitive conflict" or dissonance when they encounter the differences between a real robot and their preexisting notions about robots, e.g., from science fiction and the news media [28]. This tension, which Levin et al. operationalize as self-reported difficulty answering questions about robots, could indicate when people are reconsidering or changing their mental model [28]. The work reported in this article did not attempt to quantify differences between mental models over time or between participants; instead, the focus was on *how* models are initialized and then updated.

3 RELATED WORK

3.1 Mental Models Outside HRI

Mental models are studied in several other fields. Cognitive scientists study mental models as part of human cognitive architecture [37]. Mental models of engineered systems are studied by human factors and ergonomics researchers (see Norman [34] for a conceptual introduction) and mental models of software and other computer systems are studied by human-computer interaction (HCI) researchers [37]. There has been relatively little research on mental models of robots—a summary is given in Section 3.2 below.

A variety of theories about mental models are described in the literature. For example, in a review of research on mental models in HCI, Payne [37] lists six different senses in which the concept of a "mental model" has been used by researchers, each leading to different formulations and uses. He also talks about how mental models can be fragmentary: a person might have models for several parts of a system, but not a unified model of the whole. A mental model could be used to plan a sequence of actions using "mental simulation"—when used for this purpose, it is called a "surrogate" model, since it is standing in for the real system [37]. Moray [32] talks about how "mapping theory" and "lattice theory" have been used to describe people's models of systems at different levels of abstraction and detail. It is not known how well these theories would describe mental models of robots. In this study, a naive approach is used with only one assumption about mental model structure: that *people include the robot's perceptual and cognitive abilities in their mental models* to explain its (changing) behavior. This work did not attempt to discover the model structure used by the study participants, but rather focused on how they used their observations along with prior knowledge to initialize and update their models.

3.2 Mental Models of Robots

3.2.1 Expectations. Some of the prior work on mental models in HRI is related to people's expectations about a robot's abilities and how these expectations can be influenced by first impressions. For example, one study [25] found evidence that someone's expectation of what the robot *knows* is influenced by information about where it is from—e.g., a robot from New York City is expected to know about New York City landmarks. Other work, focused specifically on humanoid robots, looks at differences in estimates of a robot's knowledge based on its head shape [39] and apparent gender [40].

Several groups have called attention to the mistakes people make when inferring robots' abilities from their actions [3, 23, 47]. One study [3] found evidence that manipulating a robot's apparent speech capabilities can change someone's perception of its physical capabilities even though the two are often unconnected. Richards and Smart [47] have argued that this will be especially true for humanoid robots, and early experimental results seem to agree [23].

The importance of the cues people use to guess sensing capabilities was highlighted by a study of privacy concerns about "Snackbot" [24]. Very few of the people who were interviewed could identify what sorts of sensors it had—they were especially surprised to learn about the omnidirectional camera—or guessed that it was recording audio and video.

This study also looks at the expectations people have about a robot from pre-existing stereotypes or from cues given by the robot. Beyond this, our work is also long-term, qualitative, and conducted outside of a laboratory setting. This combination of study characteristics is rare in prior work.

3.2.2 Experimentation. There have been several studies of how first-time users of a robot determine what it can sense and what it responds to. Three different studies report that participants experimented with the robot by touching it, talking to it, waving, or snapping their fingers [31, 36, 43]. To allow participants the opportunity to experiment in these ways, care was taken in our study to avoid telling participants how the robot works or making the sensors too visible.

3.2.3 Mental Models of Robotic Furniture. The design of the "mobile shoe rack" in this study was inspired by several reports of robotic furniture in the literature [31, 46, 48, 53]. For example, researchers have built a robotic sofa that moves, interacts with a potential sitter, and even exhibits artificial personalities [48]. When forming a mental model of this robotic sofa, users likely borrow as much as possible from the obvious analog object—a non-robotic sofa—but this might only tell them that it can be sat upon, probably on the cushions as usual. Borrowing these aspects will be insufficient, however, for interpreting the robotic sofa's motion and inferring its sensing and decision-making capabilities. Compare this to a robot like the Aibo, which is designed so that its behaviors all have clear analogs in typical interactions between pet dogs and their owners. For this study it was deemed more interesting to make people think harder and be more creative to find non-obvious analogies or form custom models to explain the robot's behaviors.

Some people have attributed sophisticated abilities to robotic furniture. See, e.g., the report that "people [created] mental models of the [robot] trash barrel as having intentions and desires" [53]. One person treated it like a dog. This study was the first to follow robotic furniture users over many interactions to see how they develop these mental models of the robot.

3.3 Long-term HRI Studies

Some studies have deployed robots for longer periods of time to measure how interactions evolve. Sung et al. [50] in 2009 listed eight longitudinal HRI studies: two each in offices, schools, hospitals, and the home. Four years later, Leite et al. [26] published a survey of long-term studies of *social* robots in particular—they listed 23 total studies: five in health care and therapy, eight in education, five in work environments and public places, and five in home environments.

Some long-term studies use ethnographic methods. For instance, Fink et al. [9] placed a Roomba in nine Swiss households for six months; their methods included home visits, a home tour, qualitative interviews, and cleaning diaries. They included additional activities in the interviews like Bubble Talk, the Day Reconstruction Method, drawing, and tinkering [9]. In contrast, this study was designed around short, 10–20 min interviews to minimize the burden on our participants' schedules.

In some long-term studies the robot's behaviors change over time. For example, in one study [15] the DragonBot personalized its actions to individual children over a two-month evaluation. In this study, we opted for more experimental control by keeping the robot's behavioral policy the same across all participants. Instead, all participants experienced two abrupt changes in the robot's behavior that were intended to resemble unannounced software updates.

3.4 Long-term HRI Studies about Mental Models

There are very few published reports of long-term studies that focus on changes in people's mental models of robots. One example is a study of "agent migration" cues that looks at some complex intervention scenarios [21]. Another example is the study by Stubbs et al. [49], who conducted multiple interviews of museum employees who interacted every day with the educational robot PER. Stubbs et al. perform a quantitative analysis to quantify changes in participants' mental models. Conversely, in this study a qualitative analysis is performed to study *how* (i.e., by what *process*) people update their mental models of the robot using the things they notice during interactions. Much like Stubbs et al., we focus on "unmediated" interactions in which participants are not given instructions about how to interact with the robot.

4 STUDY GOALS AND APPROACH

This section begins by presenting the rationale for choosing a qualitative, hypothesis-generating study design. After that, three more subsections describe the goals and approach of this study, which are also summarized in Table 1.

4.1 Rationale for Qualitative, Hypothesis-Generating Study Design

Most HRI studies up until now take a hypothesis-testing approach. They pose a particular hypothesis about certain variables, which are first operationalized and then measured to test whether the data make the hypothesis seem unlikely. While this is undoubtedly the proper approach when comparing or extending existing theories (e.g., from human-human interactions to human-robot interactions), it is less applicable to exploratory studies. Mental model development in human-robot interaction is relatively unexplored, so it is unclear how well existing theories like the ones reviewed in Section 3.1 will fit.

It was an intentional decision to avoid committing to any particular theoretical framework for this study. We wanted to explore mental model formation about robots with open minds and to remove temptations to fit the data into any particular framework. The goal was to let the specific research questions, variables to measure, and (eventually) theories arise from a systematic analysis of data collected in a natural setting.

Thus, a *hypothesis-generating* approach was used, because the goal, besides discovering interesting phenomena, was to produce data-driven research questions for targeted testing in future work. Data were generated by asking people for detailed descriptions of their experiences with the robot. The data analysis was *qualitative*; for example, no attempt was made to quantify differences

Table 1. Summary of Goals for This Work

GENERAL RESEARCH GOAL
Study how people form mental models over time about a robot's perceptual capabilities, rules of behavior, and communication with other humans. For example, ...
… what expectations they have before their first interaction with the robot.
… how their preliminary models are formed during the first few interactions.
… how these models are changed over time as they experience more.
… how they cope with sudden changes to the robot's rules of behavior.

SPECIFIC STUDY GOALS …	… and the methods we chose to achieve them
Methodology	
○ Collect periodic descriptions of mental models without biasing future measurements	• nonintrusive video • indirect interview questions
○ Report should describe a process, not quantify effects	• qualitative (case study) analysis
Robot	
○ Non-anthropomorphic robot with a minimum of sensor cues	• a "mobile shoe rack" • partially hidden webcam
○ Participants form their own mental models	• minimal instructions or description of the robot by the study team
○ Realistic use case so people play along	• storing shoes during yoga class
○ Full control of (changes in) robot behaviors	• Wizard of Oz technique
Setting	
○ "In the Wild"	• in the hallway outside a yoga classroom
○ The same people come there regularly	• participants had registered for a biweekly class
○ Wide age range of adults	• university faculty and staff

The general research goal is presented first with some example topics of interest, followed by the specific study goals. Each specific study goal is matched with the aspect of the study methods that fulfills it. See Sections 4 and 5 for details.

in mental models over time or between participants. In fact, no complete picture of a mental model is presented; instead, the processes by which the models change are captured and described.

Qualitative analysis has several other advantages for this study besides being well-suited for early, exploratory research. Qualitative methods tend to focus on uncovering a *process* instead of accounting for the *variance* of certain variables [30]. Since qualitative methods emphasize *how* two events are linked and not the magnitude of the relationship, they are well-suited for uncovering the causal chain and making recommendations for how to improve something (e.g., a human-robot interaction). Finally, qualitative methods tend to focus more on individual people and their specific contexts instead of on differences between group averages. This can help researchers identify new, unexpected phenomena.

4.2 Venue and Robot Application

To improve experimental realism it was decided to do a robot deployment "in the wild" (i.e., outside the lab) for multiple weeks. A classroom setting was chosen to allow observation of the same group of people at least once per week and to minimize the risk of missing data due to participant absences. To obtain a wide age range of participants, faculty, and staff courses were preferred over undergraduate courses.

It was also thought that the robot should have a useful role in our chosen setting, as pilot studies revealed that some participants (and spaceowners) are no longer happy to "play along" with the experiment when the robot does a useless task, does its task very poorly, or makes life more difficult. These participants seemed to focus their comments on their criticisms and would become resentful towards or dismissive of the robot. Avoiding this became a goal of the study: It is preferable for participants to be relatively willing to interact with the robot so that their mental models would develop over time.

4.3 Robot Appearance and Behavior

The robot for this study was designed to have a novel appearance and function so that people's existing mental models (e.g., for a pet dog or Roomba vacuum) would not satisfactorily account for the robot's behaviors without some changes. This was intended to force people to undergo a more lengthy and complex learning process. The robot was also designed to minimize obvious clues about the robot's sensing abilities and connectedness to humans, including anthropomorphic features (e.g., eyes, ears, a head) and easily identifiable sensors. The robot was kept wireless to make it ambiguous whether it could send sensor data to a human. All these obvious explanations for the robot's behaviors were eliminated to force people to either use their preconceptions or to make inferences from what they observe—both of which are of interest for this study.

The robot's behaviors were designed to be deterministic and not stochastic so the robot always did the same thing given the same situation. They were also designed to be realistic for a present day autonomous robot. This was intended to make the robot seem more "robotic" and to make it less obvious that the robot was being remotely operated.

4.4 Measurement Goals

Semi-structured interviews were chosen for two reasons. First, using self reports allowed for the examination of mental phenomena like thoughts and beliefs that would be difficult to extract from simply observing participant behavior. Second, semi-structured interviews were chosen instead of questionnaires to allow interviewers to ask follow-up questions to interesting responses.

Interviewers were careful to avoid asking interviewees any direct questions about the robot's capabilities to avoid prompting them to pay an unnatural amount of attention to the robot, or to be artificially motivated to try to figure out how it worked. For example, if a participant interacts with the robot multiple times without considering whether it can hear, then it is undesirable to spoil our ability to record that phenomenon by asking about hearing before the participant first thinks about it. Instead, general questions were asked about the participant's interactions with the robot—what they did, saw, heard, thought, and felt, as well as their impressions of it.

Interviewers were also careful not to suggest certain answers to the interviewees or lead them down a certain mental path. Open ended questions were preferred, such as, "Tell me about your experiences with the mobile shoe rack today," and to use the interviewee's words when asking for more details: "You mentioned that it has a 'motion sensor'—could you talk more about that?"

5 METHODS

A "mobile shoe rack" was deployed in the hallway outside of a yoga classroom on a university campus during class time.

5.1 Venue and Population

The yoga class chosen for this study is attended by faculty, staff, and graduate students. People who attend must remove their shoes before entering the classroom, but there is no good place to put their shoes—they cause a mess in the classroom but are not allowed in the hallway. A robotic shoe rack gives people an approved place to leave their shoes and can even drive up to people to collect them. Its movement was restricted to a taped-off area around a bench where people take off their shoes (see Figure 1). It quickly became apparent that students rarely talk to each other, suggesting that the focus on individual-level phenomena was appropriate.

Participants were recruited for interviews by e-mail and also via in-person announcements during class. Only six potential interviewees volunteered, so all six were accepted. Additionally, anybody who stepped inside the taped-off area was captured on video and audio, including some students from other classes; this was explained by signs posted on the walls. The participating class ran from 12:00 to 12:50 p.m. on Mondays and Wednesdays.

5.2 Video and Audio Recording

A video camera was mounted high up on the wall opposite the taped-off bench area where the robot was deployed. The camera was positioned so its field of view ended at the tape border and extended high enough to see participants' faces (see Figure 1). The camera was configured to record for the entire time when the mobile shoe rack was inside the taped-off area. The camera was pointed up when it was not recording so that it could not see anyone.

5.3 The Robot

5.3.1 Appearance. The mobile shoe rack was built on the Turtlebot 2 platform. Figure 2 shows how it looked to participants. The top surface of each shelf was covered by paper with pale orange shoeprints painted on it to suggest that it is intended for shoes. Besides this, the robot had no nametag or instructions posted on it. It was identified as a shoe rack by two signs at the edges of the taped-off area as well as during an announcement by author Rueben in the yoga class and in e-mails to people with offices in the hallway.

5.3.2 Networking. The robot was teleoperated using the Wizard of Oz technique via a Wi-Fi connection. Network connection dropouts occasionally occurred—short (approximately 1–15 s) periods when the wizard either could not drive or could not see any video from the robot. During these times the robot stopped moving. This was most conspicuous in Condition #1 (see Section 5.4) when the robot happened to be driving across the taped-off area or turning around. It was less conspicuous when the robot was in the home position in Conditions #2 and #3, because the robot only moved occasionally in that situation. Sometimes this happened as often as once every 1 or 2 min.

5.3.3 Sensors. A webcam with a built-in microphone was mounted underneath the front edge of the shoe rack to provide a video feed to the wizard for driving. Audio was recorded from the microphone, but not piped to the wizard. These recordings were intended to capture conversations about the robot, but noise from the robot's motors, wheels, and from people walking on the creaky floors made it difficult to hear most conversations.

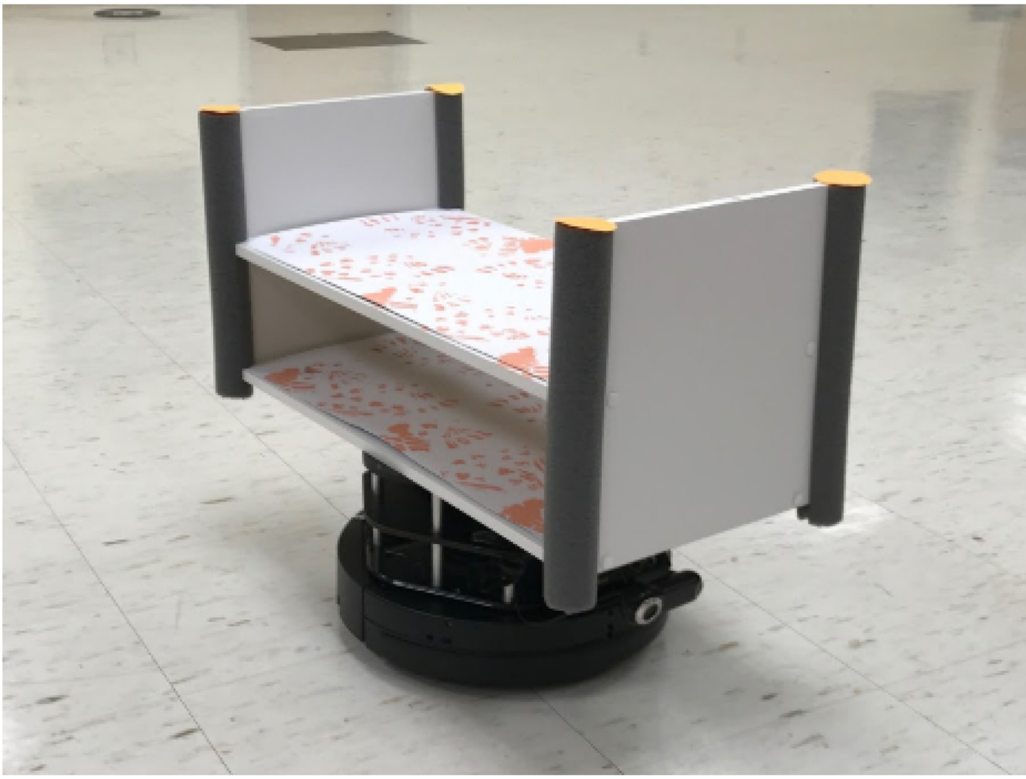


Fig. 2. Closeup of the “mobile shoe rack” robot. Webcam is visible under the right edge of the shoe rack. Foam was wrapped around the four corners of the shoe rack to prevent hurting people if the robot bumped into them (or vice versa) even though the robot drives slowly. An orange circle capped each foam tube to make it possible to automatically track the robot’s pose from the wall-mounted webcam.

5.4 (Simulated) Robot Capability Conditions

The robot’s behaviors were manipulated to produce three conditions, each two weeks in duration. Each condition consisted of a set of behavioral rules that were designed to simulate a set of sensing and processing capabilities. During each condition the robot was controlled by a human wizard according to that condition’s behavioral rules. The robot capabilities for each condition consisted of the capabilities from the previous condition plus some new ones. All sounds made by the robot were custom made by the study team using the free software Audacity. See Figure 1 for an overview of the study area.

5.4.1 Condition #1.

- **Simulated Perceptual Capabilities:** The robot can detect obstacles. It can also detect people as different than inanimate objects.
- **Sounds:** We created a “bee-boop” sound to get people’s attention and signal that the mobile shoe rack was ready to receive (or return) their shoes. The two tones were a higher pitch followed by a lower pitch in quick succession. The “car-horn” sound was based on the sound made by an impatient driver on a car horn: two medium-length blasts in quick succession.
- **Behaviors:** The robot drove repeatedly along the length of the bench. It stopped and made the “bee-boop” sound when someone was on or right in front of the bench. A detailed behavior policy is shown in Figure 3.

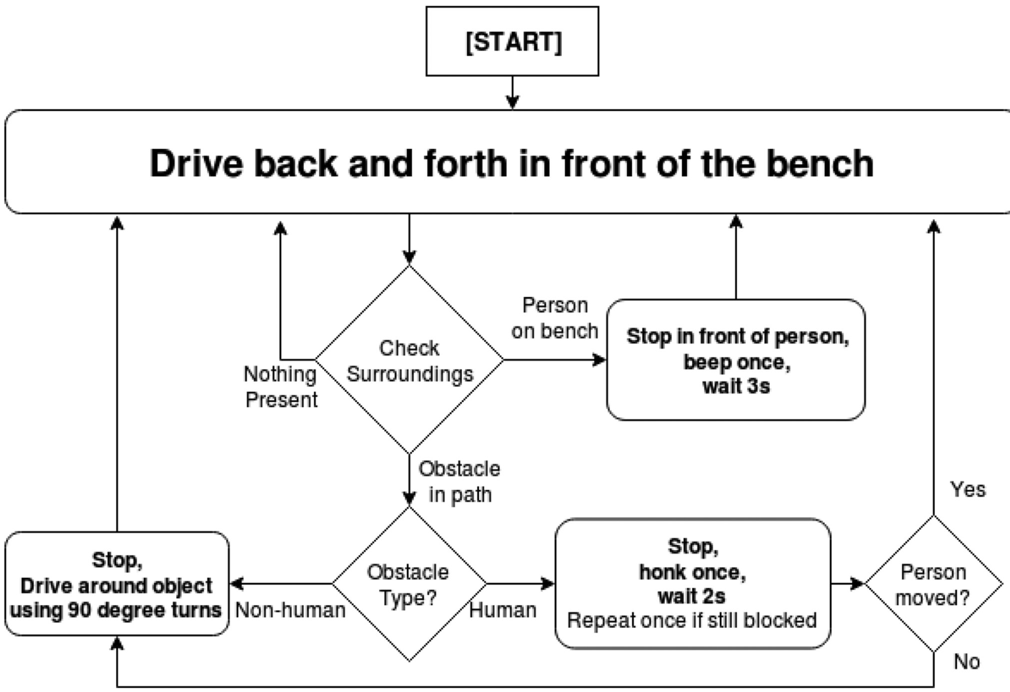


Fig. 3. Flowchart of the robot's behavior in Condition #1.

5.4.2 Condition #2.

- **Simulated Perceptual Capabilities:** The robot can do everything from Condition #1, plus distinguishing between individual people and remembering who it has visited within the entire taped-off area. It can also tell when shoes are added to or removed from the rack after solicitation via the "bee-boop" sound.
- **Sounds:** The "sad sound" was designed to make the mobile shoe rack sound disappointed. Two tones were used as for the "bee-boop" sound, but these were longer, and each fell in pitch.
- **Behaviors:** A detailed behavior policy is shown in Figure 4. Starting in this condition, the robot did not drive back and forth when there were no people to serve; instead, it waited in a "home position" in the middle of the hallway near the edge of the box that is farther away from the yoga classroom. Every 10 s or so, the robot turned slightly to one side and then to the other to "look for" people. If the wizard saw someone enter the taped-off area, then that person was selected as the robot's target and the robot would begin driving straight towards them. Also, the "sad sound" was issued if the robot reached its target but no shoes were exchanged within several seconds of playing the "bee-boop" sound.

5.4.3 Condition #3.

- **Simulated Perceptual Capabilities:** The robot can do everything from Condition #2, plus detecting when someone is making eye contact with the robot. It can also detect when its current target is in the process of taking off their shoes.
- **Sounds:** The "blabber" function was designed to make it sound like the shoe rack was talking, albeit in a robot language of electronic beeps. We created a library of 12 audio clips

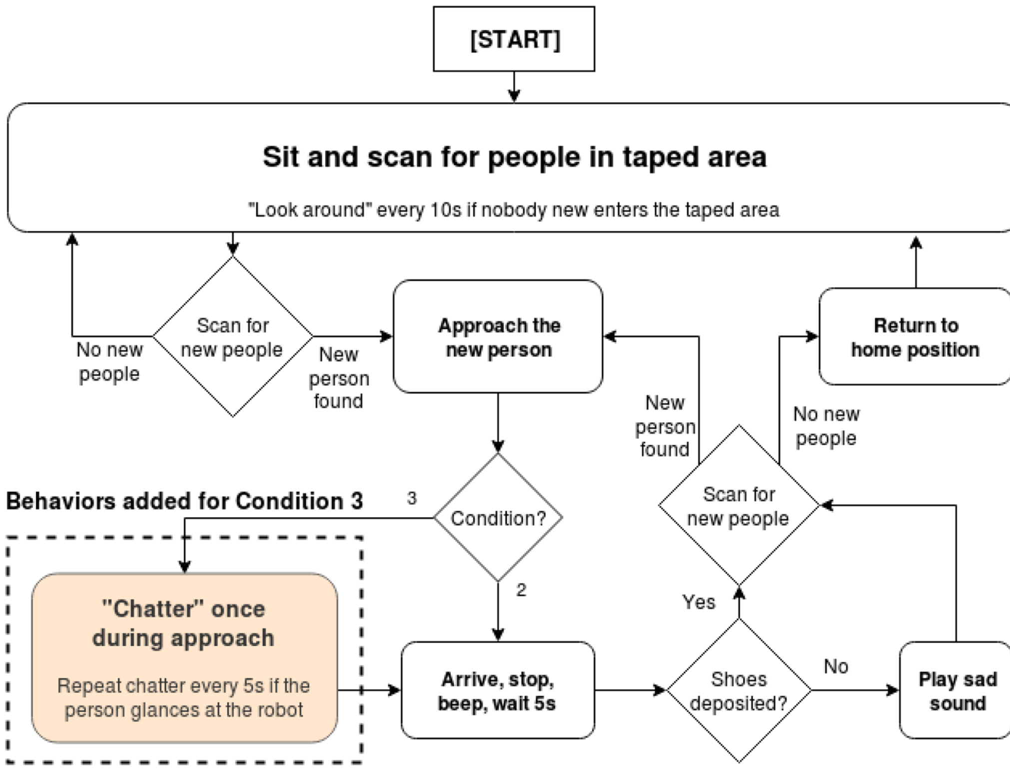


Fig. 4. Flowchart of the robot's behavior in Conditions #2 and #3. In cases where the target person left the taped-off area during the robot's approach, the robot would stop moving, play no sounds, and resume normal execution from "Scan for new people."

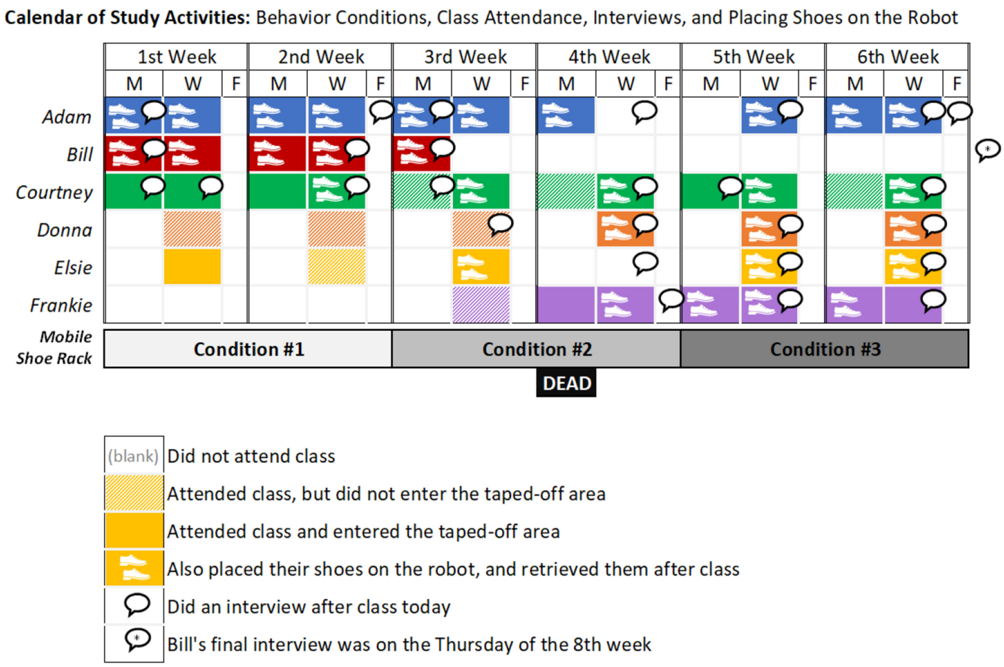
ranging from 0.4–1.1 s long that each sounded like a single word or short phrase; each "blabber" action by the robot would randomly play one of these clips. The beeps varied in length and timing to mimic the cadence of human speech.

- **Behaviors:** The robot mostly behaved the same as in Condition #2 except that as it was driving towards a target person, it did a "blabber" if that person looked at it. After a cool-down period of about 5 s, the robot would "blabber" again if the person glanced at it, or had been staring at it. This new behavior is shown in Figure 4 at bottom left.

5.5 The Wizard

The mobile shoe rack was remotely controlled by author Klow. Robot behaviors and operation were practiced at the chosen venue for four weeks prior to the main study, interacting with students from two different classes. Although robot driving became quite consistent with practice, there were occasional deviations from the behavioral protocol, including: (1) driving closer to or farther from the bench in Condition #1, (2) forgetting to play the correct sounds at the correct times, especially when attempting to coordinate action timing and path planning in Conditions #2 and #3; and (3) imprecise timings—e.g., for the 10 s between looking to either side in the latter two conditions, or for the 5 s "blabber" cool down. Error 2 in particular probably made it harder for those participants who saw it to figure out the patterns in the robot's behavior.

Table 2. Study Schedule and Diagram of Participant Attendance and Participation (i.e., Whether They Entered the Taped-off Area or Put Their Shoes on the Robot), as Well as When the Interviews Happened and When the Robot Conditions Changed



Class was held Mondays and Wednesdays, as were most interviews, though some interviews occurred on Fridays. On the fourth Monday the robot was out of battery and was placed in the "home" position by the study team (see Section 5.7). This day is labelled as "DEAD" above. Bill and Frankie were the instructors and traded places midway through the third week. All participant names are pseudonyms.

5.6 Procedure: A Typical Class Time

The mobile shoe rack was carried to the venue and placed it in the "parked" position at around 11:35 a.m. It then began driving at 11:45 a.m. by driving away from its "parked" position against the wall and beginning the behavioral protocol for whichever condition was active during that two-week period. The previous class was scheduled to end a few minutes later at 11:50 a.m., and the participating class began at 12 noon. The robot continued to drive until 12:05 p.m. to catch any students who came late. The robot then parked until 12:40 p.m. (10 min before class got out), but video and audio were still being recorded from the wall-mounted camera in case people interacted with the robot while it was parked. At 12:40 p.m. the robot began driving again to return people's shoes using the same behaviors as before class. This continued until 1:05 p.m. when the robot parked itself and an experimenter came up to collect the robot.

5.7 Timeline of Study Activities

Table 2 shows the study schedule, including the three behavioral conditions and the dates of all the interviews with each of the six participants who volunteered to be interviewed.

Study activities began with an announcement by author Rueben during the first yoga class of the term. Students were told to expect the "mobile shoe rack" before and after class, and that they could put their shoes on it if they wanted to. No study activities happened that Wednesday; trials started the following Monday and spanned six weeks (12 class times) total.

These six weeks were divided evenly into three conditions of two weeks each. Transitions between conditions were done abruptly and without warning the students—on the first day of a new condition, the robot started doing its new behaviors as soon as it started driving.

One day was an exception; Monday of the fourth Week, during Condition #2, the robot was accidentally left unplugged overnight and the battery died. It was placed in its "home position" (see Condition #2 details above) and left there, now an "immobile shoe rack," for the entire time that it was usually in the taped-off area. The participants were not informed that the robot's batteries had died.

Interviews were done about once per week. Interviews were targeted for days when interviewees would have more to say: the first day of a new condition when the robot's behavior had suddenly changed, and the last day of a condition when participants had had two weeks to observe it. Interviews did not always conform to this schedule, though, especially for "Elsie" and "Donna" (see Section 5.10 below) who only attended class—and only did interviews—on Wednesdays.

5.8 Interview Protocol

Author Rueben did all the interviews except for several that were done by author Zimmerman. Interviews were recorded on a handheld audio recorder, and interviewees were compensated $10 per interview. Interviews were semi-structured: interviewers were given a list of carefully designed questions to ask, but were also trained to give encouraging prompts and ask follow-up questions. The full interview guide is included in Appendix A.

The first part of the interview was meant to get the interviewee's account of everything they noticed about the mobile shoe rack since the previous interview. Specific details were preferred, so the interviewer asked follow-up questions to get the exact sequence of actions done by both the robot and the interviewee.

The second part of the interview asked more general questions. Some examples include, "What is your overall opinion of the mobile shoe rack?"; "If you could make it better, what would you change?"; "Describe all the shoe rack's behaviors as best you can"; and, "Would you want the mobile shoe rack in your home? Why or why not?" The interviewer also asked, "Do you have any questions about the mobile shoe rack?", but the interviewer did not answer these until after the final interview.

After the final interview, the interviewer asked direct questions about the interviewee's mental model of the mobile shoe rack. The questions were specific and covered a wide variety of capabilities: "Can it see? How far? Can it distinguish people from inanimate objects?"; "Can it hear?"; "Does it remember that you've already put your shoes on it?"; and several others. After these direct questions, the interviewer revealed that the study was focused on how they figured out the robot's capabilities. A description was given of what the robot's capabilities were in each condition. It was also revealed that the robot was driven remotely by a human.

5.9 Interview Analysis

The interview analysis followed a case study research approach [4, 30, 54]. All the transcription and analysis of the interview recordings were done by author Rueben. Each interviewee's experiences were analyzed as a separate case. First, interview responses were placed in chronological order to form a coherent story for each interviewee. Next, the analyst identified all mentions of the robot's capabilities or behaviors, or of observations or reasoning processes by which the interviewee says they learned about its capabilities or behaviors. Whenever an interviewee's response was unclear or seemed to contradict the robot's typical behaviors the video recordings were reviewed to find out what really happened. Finally, all the mentions of the mental model formation process from the six cases were reviewed together in a cross-case analysis to identify interesting questions and phenomena.

5.10 The Interviewees

All interviewees were recruited from the faculty-staff fitness yoga class that the mobile shoe rack was serving. Six people volunteered to interview with the study team, and were asked to complete a demographics survey to gauge the diversity of our sample. The interviewees' names have been changed for confidentiality:

- "Adam" is a male in his 30 s who does STEM outreach education and professional development for teachers.
- "Bill" is a male in his 60 s and the main instructor for the faculty-staff yoga class, but left for vacation after the third Monday of the study.
- "Courtney" is a female graduate student in her early 20 s studying marine resource management.
- "Donna" is a female in her 70 s who was faculty in the College of Pharmacy and then worked at Student Health Services—she only attends class on Wednesdays.
- "Elsie" is a female in her 40 s who works as a licensed psychologist for Counseling and Psychological Services—she also only comes on Wednesdays.
- "Frankie" was the substitute yoga instructor for "Bill," so her first day was the 3rd Wednesday of the study; she identifies as gender non-binary and is in her late 20 s.

Table 2 shows which days each interviewee attended class, and when the interviews actually happened.

6 OVERVIEW OF INTERVIEW DATA

This section is an introduction to the relatively large set of interview data collected during this study. Some metadata are provided about interview duration and per-participant interview frequency, and then some characteristics of the interviewees' responses are described that will set the stage for the findings presented in the next section.

6.1 Interview Data Collected

In total, 28 interviews were conducted (see Table 2) with six participants, whose names have been changed for confidentiality. "Adam" and "Courtney" each did seven interviews, since they were present for the whole study; "Donna" (who did four) and "Elsie" (who did three) volunteered partway through. "Bill" was interviewed three times before leaving on vacation and once after returning (after the study ended); "Frankie," who replaced him, did three interviews. The final interviews were longer than the non-final ones. All of the non-final interviews lasted around 5–20 min except for one with Adam that was almost 30 min long. Final interviews lasted longer because of the direct questions about mental models as well as the debrief; most were around 30–40 min, although Adam's was over an hour long.

6.2 General Observations About What People Said

6.2.1 Interviewees Did Not Observe the Same Robot Behaviors. Each interviewee had a unique set of experiences with the mobile shoe rack. Figure 5 shows how several factors contributed to this. First, the robot's behaviors could change from day to day within a condition, because it reacted to people around it. For example, in Condition #1 someone needs to be sitting on the bench for the robot to stop and beep at them; if nobody is on the bench, the robot will never stop or make the "bee-boop" sound. Second, interviewees did not always attend class, and spent different amounts of

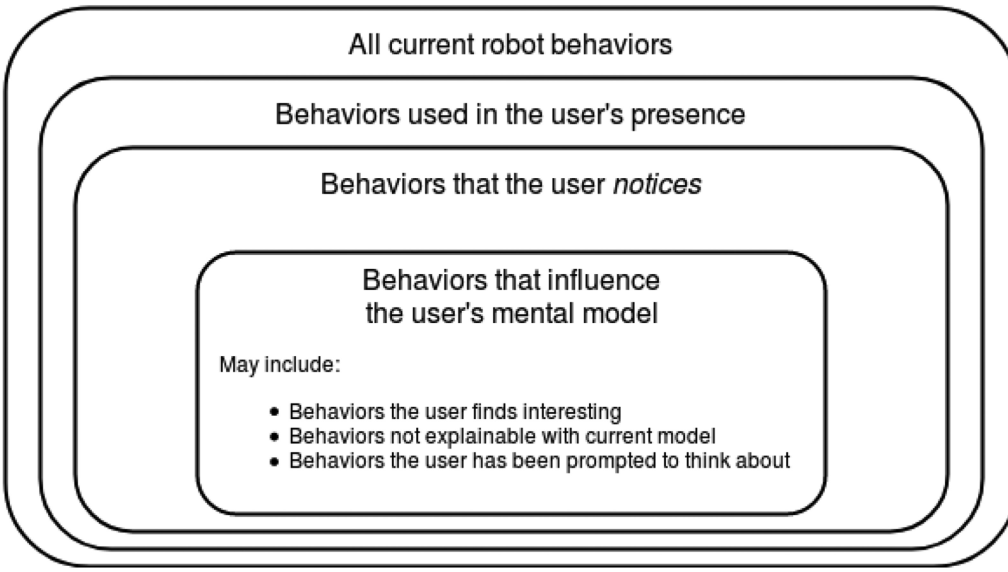


Fig. 5. A proposed Euler diagram of how participants come to observe only a subset of the robot behaviors being expressed, which influences the mental model formation process. Note the disconnect between the intended behaviors programmed by designers and the observed behaviors that can influence users' mental models of the robot, discussed further in Sections 6.2.1 and 6.2.2.

time observing or interacting with the robot (see Table 2). Some interviewees would walk quickly past the taped-off area and barely glance at the robot; others would pause just long enough to take off their shoes. For those who put their shoes on the mobile shoe rack, some would sit on the bench watching the robot for a few minutes whereas others would rush into the classroom. Bill and Frankie often arrived very early and others sometimes arrived late—in both cases, they were often the only person in the hallway with the robot. Other interviewees arrived during the transition time between classes when the hallway was crowded; they saw the robot interact with several other people.

This second factor—the amount of time and attention people gave to observing the robot—was especially relevant for longer behaviors. For example, Adam on the sixth Wednesday mentioned that he noticed (apparently for the first time) that the robot visited two people in a row without returning to its home position in between. The robot had actually had this capability for four weeks by that time, but to observe it Adam had to be paying at least periodic attention to the robot for a relatively long period of time in addition to being there with at least one other person whom it had not yet visited. Another example is the robot's capability in Conditions #2 and #3 to recognize you as an individual—to get evidence of this, you have to notice that the robot visits you once but then wait long enough to see that it does not visit you again. These sorts of behaviors were less likely to be observed (fully) by our interviewees compared to shorter, simpler behaviors.

Due to these factors each interviewee only got a sample of the full set of possible behaviors for the current condition on a given day. It was therefore difficult to compare their mental model development processes to each other directly. Instead, the data from each interviewee were analyzed as an independent case study to identify interesting phenomena, and then a cross-case analysis was performed across all six of these cases to put together the findings presented in the next section.

6.2.2 Interviewees Varied Widely in Motivation to Understand the Robot. Adam had the highest motivation of the six interviewees to figure out how the robot worked and the purpose of the study. This started with the in-class announcement about the study before he had even seen the robot. By the interview on the fourth Wednesday he said, "Every day I try to figure out what's going on with it," and later he referred to the "entertainment value" of experiencing all the behavior changes—"It's like a puzzle and you wanna figure it out." This high motivation probably increased the amount of attention that Adam paid to the robot, how much he thought about its behaviors, and how often he reconsidered or revised his mental model of it.

Some of the other interviewees avoided or even ignored the mobile shoe rack at the beginning of the study. In particular, Courtney and Donna did not put their shoes on the robot for the first three times they saw it, and Elsie and Frankie for the first two times (see Table 2). Donna did not even enter the taped-off area during those first three class times. During her first interview on the third Wednesday she expressed a desire to observe the robot more before answering some of the interviewer's questions and said she would "be more observant" the next time, perhaps explaining her increased attention and participation from that point forward. In fact, all six interviewees were at least *using* the mobile shoe rack regularly by the end of the study. It seems likely that this is due to feeling obligated or incentivized to have something to say at the next interview.

7 FINDINGS FROM THE CROSS-CASE ANALYSIS

The interviews were reviewed for each of the six interviewees to identify everything they said about the robot's behaviors and capabilities. Interesting phenomena from these six cases were then aggregated and compared. Research questions and phenomena that warrant further attention have been organized into six topics and are presented in this section. Topics were chosen for different reasons: Section 7.1 highlights how the interviewees' model formation processes were *different*, Section 7.2 shows the *variety* of ways that the interviewees reasoned about what they experienced, and Sections 7.3–7.6 each assemble a group of phenomena that were observed to be *similar* across most or all of the interviewees. Table 3 should help the reader navigate all our observations from the interviews, including those presented in the previous section.

7.1 Variability of the Rate at which the Predictiveness of the Mental Model Was Improved

Over time the interviewees noticed the mobile shoe rack's different behaviors and tried to update their mental models to better explain them. The updated models accounted for more of the robot's behaviors and predicted the occurrence of these behaviors more accurately. The amount of time it took for the interviewees to be able to predict the robot's behaviors with high accuracy might not generalize very far outside the particular setting and robot use case of this study. Interestingly, though, considerable variance was observed between the six interviewees. Some people noticed and had developed explanations for new robot behaviors after one or two days of experiencing a new condition. For example, Adam and Bill understood by the end of the 1st Monday that the robot stops in front of each person and waits "for a second" (Bill's words) before moving on. For other people it took more time before their model was very predictive—on the same day, Courtney said it stopped seemingly "at random" and did not figure out that it was stopping for *people* until the following Wednesday. Still others like Elsie had not accounted for some of the main elements of the robot's behavior in their final model at the end of the study.

Several factors could account for this variance. For example, differences between the interviewees in motivation to figure out the robot have been highlighted in Section 6.2.2. Adam seemed to have the highest motivation, followed by Bill, and Courtney seemed to have the lowest (although

Table 3. A Guide to the Observations Reported in the Interview Data Overview (Section 6) and Cross-case Analysis Results (Section 7)

6.2 General Observations about What People Said
6.2.1 Interviewees did not observe the same robot behaviors
6.2.2 Interviewees varied widely in motivation to understand the robot
7 Findings from the Cross-Case Analysis
7.1 Variability of the Rate at which the Predictiveness of the Mental Model was Improved
7.2 Reasoning about Evidence and Hypothetical Situations
7.2.1 Drawing conclusions
7.2.2 Deciding not to draw a conclusion
7.2.3 Forming hypotheses and hypotheticals
7.2.4 Discussion
7.3 Comparing the Robot to Humans, Animals, and Other Devices
7.3.1 Comparisons and Associations
7.3.2 Borrowing (pieces of) other Mental Models
7.3.3 Borrowing from Existing Mental Models of "Robots"
7.3.4 Borrowing Structure vs. Content
7.3.5 Intentionally Projecting Attributes onto the Robot.
7.4 Attributing Sensing Capabilities without Visible Sensors
7.5 Judging whether the Robot is Autonomous or Teleoperated
7.5.1 Why did not anybody suspect a teleoperator?
7.6 Experimenting with the Robot

it should be noted that those with low motivation generally started participating more after their first few interviews). This motivation might drive people to improve the predictive accuracy of their mental models more quickly by increasing attention, thoughts about the robot, and how often the mental model is reconsidered or revised.

Interviewees' experiences of the robot also varied considerably (see Section 6.2.1 and Figure 5), and could therefore be another factor influencing the rate at which the predictiveness of their mental models improved. Differences in attendance and participation are presented in Table 2. Two participants—Donna and Elsie—only attended class on Wednesdays. Bill the instructor left partway through the study and was replaced by a substitute, Frankie—both missed entire conditions. Some of the interviewees reported ignoring the robot as they hurriedly walked to and from class on some days, whereas others sat on the bench for several minutes watching it. It seems reasonable to assume that people update their mental models of the robot's behaviors based on which behaviors they actually *experience*, which in this study was not always everything in the behavioral protocol. Longer, more complex behaviors are probably more vulnerable to being missed, because they require a person to observe them for a longer time period to understand what is happening. Interaction designers need to understand this when they are thinking about how to teach their users about these behaviors—a longer demonstration would be needed, or else a verbal or pictorial description summarizing it.

Other factors that could account for different rates of improvement of model predictiveness could include: memory, ability to reason about evidence, attention to detail, and other individual differences. Future work could use a framework such as the one presented by Funder [11] to ensure that no important sources of variance are overlooked. Funder writes that the accuracy of personality judgments in human-human interactions depends on four factors. Applied to mental model formation about a robot, they might be stated as:

(1) "the **relevance** of [the robot's] behavioral cues to [elements of the human's mental model of the robot],"
(2) "the extent to which these cues are **available** to observation,"
(3) "the extent to which these cues are **detected**" by the human, and
(4) "the way in which these cues are **used**" in the human's mental model formation process [11].

Factors 1 (relevance) and 2 (availability) emphasize the roles of the robot and the environment, whereas factors 3 (detection) and 4 (utilization) recognize differences between humans as perceivers. Figure 5 offers a graphical depiction of how factors 2 (availability) and 3 (detection) could have figured into the interviewees' experiences, whereas this section has mostly discussed individual differences between the interviewees, which influence factors 3 and 4 (detection and utilization). The next section focuses specifically on factor 4—how interviewees *used* the robot behaviors they noticed to update their mental models.

7.2 Reasoning about Evidence and Hypothetical Situations

Several methods were observed that the interviewees used to reason about evidence—either to draw a conclusion or to decide the evidence is inconclusive—or to reason hypothetically about events that did not happen but might suggest different (or clearer) conclusions.

7.2.1 Drawing Conclusions. The interviewees made several different types of inferences—some correct, some incorrect—from the robot behaviors they saw.

Simple inference: A popular example: people would often notice how many shoes were on the robot and make inferences about how many people used it.

Correlation: Some rather impressive connections were made between the robot's behaviors and the things that people thought were causing them. For example, Courtney on the fifth Monday reasoned that she could not hear the beeps from inside class, because there was only one person on the bench when she emerged, whereas in the past there were more, and she could hear it.

False correlation: On the sixth Wednesday before class Frankie walked out of the yoga classroom and stood barely inside the zone watching the robot, "Just to see if the robot was maybe on a timer." After a few seconds of waiting, the time came for the robot to start driving, and it came up to Frankie. She said "it noticed me," incorrectly believing that her presence had triggered it.

Offering explanations for unexpected behaviors: Early in a new condition and especially at the very beginning of the study the interviewees did not know what to expect, and were trying to find explanations for the robot's behaviors. For example, on the fourth Wednesday when the robot moved away from Courtney before she could put her shoes on it she suggests as an explanation the fact that there was "someone else in the box"—maybe it was trying to visit them. She may have chosen this explanation, because it fit with a belief she was holding about how the robot visits people. Also, that other person's presence may have been more salient to her than other details like the shoes on the floor.

Interestingly, almost nobody talked as if the robot makes mistakes—i.e., via sensor malfunctions or computation errors. A rare exception was Bill's frightened thought that the robot may have "malfunctioned" when it first drove directly towards him on the 3rd Monday. Perhaps existing theories from organizational studies about *sensemaking* could be applied to these sorts of bewildered responses [51]. A few of the interviewees did talk about *randomness* in the robot's behaviors—once by Courtney on the first Wednesday when she had no better explanation the robot's movements and again by Adam on the fifth Wednesday to describe how its behaviors (in Condition #3) were less predictable and more "humanlike."

Eliminating explanations: When Adam noticed the new behaviors on the 3rd Monday he quickly eliminated the possibility that they were triggered by the fact that he was the only person in the area: "the situation didn't seem unique, I've come in late before...so it wasn't like, just because nobody else was there that's what it was doing."

7.2.2 Deciding not *to Draw a Conclusion.* There were also several reasons that the interviewees gave for *not* drawing any strong conclusions from an observation.

Reasoning that the evidence may not help one decide between potential conclusions. Donna in her final interview mentioned one piece of evidence that was not helpful to her. When asked whether the robot responded when she said "muchas gracias" to it, she said, "No, it...it paused a tiny bit longer, but I don't know if it would have paused had I not said that." That is, the slightly longer pause did not help her decide whether the robot responds to speech, because it might have happened without it.

Considering the fact that their memory and attention are limited: Several times somebody noticed a behavior for the first time, but reasoned that it might not be a *new* behavior—it could have been happening the whole time but they simply did not notice it.

For example, Courtney on the fourth Wednesday admitted that she was not paying as much attention at the beginning of the study as she is now, and is unsure whether behaviors like "the two different beeps" (i.e., the addition of the sad sound in Condition #2) are actually "a new thing" or if she did not notice them before. Later in the interview she even doubts herself that there really are two different beeps.

Also, when Donna finally notices the orange circles on the corners of the shoe rack on the fifth Wednesday she considers the possibility that they were actually there the whole time (which they were): "...I didn't remember that it had little orange corners prior, so I don't know, maybe I'm just getting more observant."

Waiting for more (informative) evidence: Donna on the 3rd Wednesday chooses not to answer several of the interview questions, because "I haven't had enough interaction" with it and "I will want to have some more time to engage what it does and so forth" before answering. She even reasons about the *informativeness* of different robot actions and thereby shows an understanding of statistical significance during her final interview: when asked if the robot can recognize specific people, Donna reasons that one way to tell would be to see whether the robot always turns so her pair of shoes is easy for her to grab. Since the robot only has two sides, though, this would happen 50% of the time even if the robot were guessing randomly. She therefore reasons that, although the robot did present her the correct side today it may have been "coincidental" and that she would need more interactions to test this properly. It is worth repeating that Donna is a retired university faculty member.

7.2.3 Forming Hypotheses and Hypotheticals. When interviewees found the robot's actual behaviors ambiguous they often brainstormed hypothetical situations that would be more helpful for answering the interviewer's question. These were essentially experiments—scenarios in which the robot's response would help them decide whether, e.g., the robot has a certain capability.

Reasoning about what it would do in an edge case: Donna in her final interview: "Well, so far as I know it's got an ongoing energy source. It's not like if the lights go out...[that the robot will turn off]."

In the same interview Donna suggests testing the robot's idea of the boundary between person and object by introducing "a service animal" to the situation.

Another example occurs in Bill's final interview. He was unsure about one of the questions: "I never knew if there was something else there that was an object if it would treat it the same as a person in terms of stopping in front of it." He proposes putting a mannequin ("dummy") in front of the robot to test whether it can "know that it's not a breathing, living thing."

Counterfactual evidence. Some of these hypotheticals were counterfactuals—i.e., things that did *not* happen. The logic for these was of the form, *if <this> had happened instead, then I would have concluded <that>*.

For example, Bill on the second Wednesday says, "If it knew who I was through face recognition or something, it'd say 'Oh! [These are] his shoes, so I'll bring him his shoes.'"

Reasoning about what would be observable in different situations. For example, Donna on the fifth Wednesday notices the painted shoeprints on the shoe rack's shelves: "...today I was able to see that there are actually some footprints..., and I didn't observe that before, because it was fully loaded."

7.2.4 Discussion. It is noteworthy how much reasoning people do and how sophisticated some of it is. Anecdotally, the interviewer felt that much of this reasoning was done *ad hoc* during the interviews and may not have happened without some motivation to reflect on why and how the robot did the actions that they observed. If this is true, then it suggests that the reasoning that leads to mental model adjustments does not necessarily occur immediately after observing the robot's behavior. Instead, it might be prompted later when the need arises to explain those observations or use them to produce a more accurate or complete mental model. Understanding what prompts users to reason about their observations and update their mental models could help designers encourage these updates to speed the users' learning process.

Interviewees also seemed to use reasoning to compensate for the fact that they often failed to notice the robot's behaviors or forgot about the behaviors they did notice. Trying to make confident conclusions from limited data exposes their mental model formation process to well-documented heuristics and biases like confirmation bias and self-report biases like the one reported by Nisbett and Wilson [33]. In general, the way people assign causes to events they observe (such as robot behaviors) is described by theories of attribution (e.g., by Jones [17], Ch. 3). Future research could study how well these theories also apply to user mental model formation about robots, as there is a large body of work from which human-robot interaction researchers could benefit.

7.3 Comparing the Robot to Humans, Animals, and Other Devices

7.3.1 Comparisons and Associations. Interviewees often mentioned other things when talking about the robot. Most of these mentions were simply comparisons or associations. The most common comparison—made by Adam, Bill, Courtney, and Elsie—was to the Roomba robot. Early on, Adam compares it to a pet: "You feel like you're almost calling it like you're calling a dog or something, like, 'Come here shoe rack, come on!'" Frankie compared the blabber sounds to "songbird noises," and Adam called them "R2-D2-type sounds." Frankie even made a rather abstract comparison: "I want more of a Pee Wee's Playhouse vibe out of it, and I'm more just getting [baggage carousel] at the airport with bags on it." Some of these comparisons may have *only* been comparisons—i.e., metaphors used to describe something about the mobile shoe rack.

It did not seem like the comparisons the interviewees made to the mobile shoe rack were random. They seemed to gravitate toward comparing it to either a pet dog or R2-D2. Which aspects of the mobile shoe rack (or the environment, or the scenario) caused them to choose those objects of comparison? Did a dog come to mind because the robot was subservient?... social? ... of about the right size? It will be important to predict which objects (e.g., a dog, a droid) they will choose when making comparisons to the extent that these comparisons influence the mental model formation process.

7.3.2 Borrowing (Pieces of) Other Mental Models. More significantly, though, it appears that the interviewees sometimes borrowed from their mental models of these other entities to help them build a mental model of the robot. This is in line with the findings of Kiesler [19] and indicates that this borrowing can occur even if the robot does not clearly look like a human or animal. Borrowing from other mental models could influence the associations people make between different attributes and therefore the assumptions they make and questions they ask about a new system.

It can be difficult to tell when this borrowing is happening in addition to a simple comparison. For example, Elsie on the fifth Wednesday makes a comparison—"It reminded me of some of the sounds that my toy R2-D2 makes"—but then says, "it was kind of a fun little droid." This might have been just a metaphor, but she may have also borrowed something from her model of a "droid" as she was building a model of the mobile shoe rack.

Sometimes borrowing is more obvious. For example, Adam said on the very first day that the shoe rack appears to be on a Roomba. Here he believes that the Roomba is actually a component of the system he is trying to understand, so he might have guessed that his model for what the Roomba can do would be a very good model for what the base of the mobile shoe rack can do.

Later, Adam demonstrates some inference or prediction that resulted from using part of another mental model. During his interview on the fourth Wednesday he described a time earlier in the study when he watched members of the CHARISMA lab directed by Dr. Heather Knight driving a "Chairbot" in another building on campus. Moreover, his three-year-old daughter got to ride on one—he uses this fact to reason about how much weight the mobile shoe rack could carry: "The bottom part looks very similar to Chairbot, so if Chairbot can move 30 pounds...." He later specifies some differences between the Chairbot and the mobile shoe rack: "...the Chairbots look—the bottom of them looks similar to that thing, I don't know if they're the same, but those were all being controlled by remote control obviously, I don't [think] that there's somebody remote controlling this, I would be very surprised if that was the case." He will either have to borrow just part of the other mental model or else modify it to account for this difference.

7.3.3 Borrowing from Existing Mental Models of "Robots." All of our interviewees referred to the mobile shoe rack as "the robot" at times, so they may also have borrowed from their mental model of "robots," of certain types or classes of robots, or of specific robots they have heard about or seen. The interviewees were familiar with robots from movies, like R2-D2, as well as consumer products like the Roomba and Amazon Echo; they may have also known something about certain robotic toys (Adam has a three-year-old daughter) or robots featured in news stories that have been used to assemble cars or perform surgeries. How are all these different devices represented in their minds? Is there a mental model for all robots, or are they separated into (perhaps overlapping) subclasses? There is a wide variety to account for in such a system of models—note the differences between remote-controlled cars, dolls that can hold a conversation, assembly line robots that do repetitive tasks, cruise control and lane-keeping systems in cars, and AI personal assistants.

Another question is how these preexisting models of robots can be changed, especially since the ones held by people who do not have in-depth experience with a robotic system are probably rather inaccurate. Did people who interacted with the mobile shoe rack in this study leave with a changed mental model for "robots"? If so, then was it more accurate? Perhaps their experiences prompted them to break down the overgeneralized category "robots" into more specific subclasses: social robots vs. non-social robots, autonomous robots versus remote-controlled robots, and so on. Perhaps theories from social cognition about how we do this classification on other people would apply here.

7.3.4 Borrowing Structure vs. Content. It is important to specify what about a mental model is being borrowed: just the structure, just some of the contents that fill that structure, or both. In the previous example, perhaps Adam borrowed his *entire* mental model for a Roomba, because he believed it to actually be a component of the mobile shoe rack. In the example about Elsie it is less obvious—while it is possible that she also borrowed the typical attributes of a "droid," she may have just used the *structure* of the model as a framework.

Here, a model's structure refers to the way the pieces of information are connected. For example, most humans who can understand speech can also speak, and vice versa. Each ability strongly suggests the other one in most people's mental model of an adult human. Most people probably do *not* link those abilities in their mental models of certain other things, however; for example, a telephone system might be able to automatically play a voice recording that says, "Please hold," but most people know that it will not understand if you say something in response. The ability of an automated system to generate speech does not necessarily mean that it can also understand speech, and most people's mental models represent this lack of connection.

It would be important to know, then, whether someone is first approaching a robot using the structure from their mental model for a human or for, e.g., the "on hold" message played by a telephone system. If the robot could predict the user's choice of model structure, then it could also predict the incorrect inferences and interpretations the person is likely to make about its behavior and perhaps even choose different actions or give explanations to avoid costly mistakes.

Also, designers could use the robot's appearance and behavior to encourage users to use a certain model structure, a possibility that has been previous explored for both anthropomorphic and zoomorphic robots [19, 25, 39, 40]. For example, AIBO is designed to be doglike, so users will use elements of their mental model for a dog in at least part of their initial model for AIBO. In this study the mobile shoe rack was intentionally designed to not resemble a human or animal or anything else that moves, at least in appearance. Despite this, several of the interviewees compared their interactions with the robot to typical interactions with a pet dog. The mobile shoe rack's behaviors were not designed to resemble those of a pet dog, but it is possible that there was some accidental resemblance. Or perhaps the tendency to make these comparisons is so strong that even a tiny resemblance is enough. Future work could study how the robot's appearance or behavior could be designed to encourage users to reuse the mental model from a certain entity, particularly using robot behaviors alone.

In this study the focus has been on how people update the *content* of their mental models, such as which sensing capabilities the robot has and what rules of behavior it follows. Payne [37] says that this focus on content is more common in HCI because of the emphasis in that discipline on a user's individual actions, whereas in cognitive psychology the focus is more on model *structure* (i.e., cognitive architectures), because this gives a more general understanding that applies in many different contexts. Model structure has to do with how mental models are represented in the mind—is the robot understood in terms of abstract properties, comparisons to other entities, or in some other way? Also, do people carry around one mental model—their favorite—or do they maintain a list of possible candidates, perhaps with estimates of how likely each one is to be accurate? Answering these questions about model *structure* will be important for measuring model *content* in the future, as it will help researchers ask questions that fit the way users think instead of evoking unnatural responses by using the wrong framework.

7.3.5 Intentionally Projecting Attributes onto the Robot. When she was asked to list the robot's capabilities Frankie makes a distinction. She says, "When we talk about the robot and the things we want it to do, we assign [animal or human] characteristics to it." She often thought of it as a small dog and gave it different names like "Percy." She emphasizes the difference, though, between these assigned characteristics and what it really is: "a shelf on wheels that has a sensor on it." If people often hold both of these two sets of beliefs, then it seems crucial to distinguish between them in theory, measurement, and practice.

On the one hand, people have their mental models of what the robot can really do according to the design stance [7]. This is based on thinking about the robot as a mere mechanism—pieces of plastic and metal controlled by computers to do things according to the designs of a human software engineer.

On the other hand, people might also imagine or project additional characteristics onto the robot that they know are not really there, but are useful for interacting with the robot or predicting its behavior. This could include taking the intentional stance [6]—i.e., attributing intentions to the robot. This might just be an extra ability (e.g., pretending your dog understands what you say), a certain role or relationship (e.g., talking to the robot as if it is your butler), or even a whole new identity (e.g., when a robotic Mickey Mouse toy becomes the real Mickey Mouse). Designers often encourage users to imagine a particular personality for the robot by making the robot look and act in a way that suggests that personality. When users do this, it influences the way they interact with it—they act and talk *as if* these imagined characteristics are real.

Again, Frankie is a good example of this: "I think of it kind of like one of my small dogs, like, 'Hey, guy!'" Also, even though she knows that this "animatizing" is "just in my mind," she appears to still use her model for a dog as a template for her projected mental model of the robot: "So I guess I think of this shoe rack like my dogs...like you tell 'em you're there, and then you go sit down and wait for them to come to you, essentially." That is, she uses her dogs as a template for the shoe rack, and is therefore mapping her script for interacting with her dogs onto her interactions with the robot. Or, put in different words, she evaluates whether the robot looks and acts in a way that makes it easy for her to "think of it like a dog": "It doesn't really respond to someone sitting down in the way that an animal or something small that's alive would respond to someone sitting down." So while Frankie does have a mental model of the robot's actual capabilities and behaviors, she also wants to pretend that it is a small dog. This certainly will influence her behavior around the robot, but might also influence the way she forms her non-imaginary mental model and talks about the robot to experimenters.

Perhaps HRI researchers should begin trying to distinguish between these two parts of a user's mental model of the robot—sincere beliefs about how the robot *really* works and projections that people know are not real. A few other researchers have also advocated for more work in this direction [55, 56]. It does seem like people know when they are pretending, at least when it is brought to their attention (they may have been pretending without realizing it). Presumably you could ask a person about each part of the model separately: the real life part without any pretending, and the part with imaginary stuff added on. People know the difference between a question about an actor wearing a Mickey Mouse costume and a question about Mickey Mouse the character.

7.4 Attributing Sensing Capabilities without Visible Sensors

Bill, Donna, and Elsie came to believe that the robot could sense distance or motion even though they had not noticed any visible sensors on the robot. Perhaps they inferred this from the robot's behaviors, but if they had seen the sensor and recognized it as a video camera perhaps they would have attributed more vision capabilities to it. This suggests the importance of "sensor transparency" [42] in robot design—among other things, so that people can ask well-informed questions about their privacy before consenting to interact with a robot.

However, Adam, Courtney, and Frankie all assumed that the robot could not hear, perhaps because they never saw the robot respond to sound. The hallway was often noisy during the transition between classes, and we know of only a few instances of people talking to the robot or making noises (e.g., clapping their hands) to signal it. Even so, we cannot rule out the possibility that each of these three interviewees saw the robot fail to respond to an audible message that was clearly aimed at the robot. Another possibility is that they could account for all the behaviors we gave the robot using only vision or motion sensing, and were content with that simpler, more parsimonious model. Finally, perhaps it was because the robot does not resemble a human or animal, nor did it have anything resembling ears. Future work should work to discover why people attribute some sensing capabilities but not others to a robot, even without obvious cues to help them. In cases wherein people are unlikely to make the right attribution, robots could be designed to help. For example, the mobile shoe rack did not use the webcam's microphone for any of its behaviors, but it was still recording; if it had a display indicating the sound levels it was detecting that might remind people about the microphone.

7.5 Judging whether the Robot is Autonomous or Teleoperated

An important aspect of one's mental model of a robot is whether it is autonomous or teleoperated. Participants were divided over this in the robotic trash barrel study [31]: of the 28 people who commented on whether it was autonomous, 16 thought it was (often commenting that it was "clumsy") and the other 12 thought it was not (often describing its actions as more calculated or purposeful).

Despite our concerns that participants would realize the robot was being remotely operated, all six of the interviewees were surprised to hear during the debriefing that the robot was being operated by a human. Donna was the only person who ever suspected anything: on the fifth Wednesday she mentions that she knows "there is both visual and auditory that's being captured," and says it is possible that "somebody is actually remotely doing something" even though her best guess is that there is "some kind of sensing device on the robot itself."

The other five interviewees never suspected that there was a human operator. Some of them even got hints that could have caused them to suspect it.

Adam was the best example. He said he "never expected" it, but was also the most suspicious interviewee. He said that every time he came to class he was trying to figure out what we had changed about the robot. He actually thought he might have become "hypersensitive" to small changes in the robot's behaviors: "Am I actually picking up on things that are happening, or is my mind just doing these things?" He even develops what he calls "conspiracy theories" by the end of the study. For example, on the sixth Wednesday he noticed that the robot did not visit a certain young man sitting on the bench. He began to suspect that he had been "planted there" by the study team, and then became "more and more" suspicious of this as he noticed certain people who were "kind of out here consistently." He even admits that his experience with one of Dr. Heather Knight's Chairbots on campus should have made him suspicious, because the "the bottom of the Chairbot looks just like the bottom of this robot," he "saw people controlling them remotely," and he even asked one of them, "can it move autonomously?" and got the response "no, not yet." He admits he should have concluded that, "if they didn't do it with [the Chairbots] how could they do it with the shoe rack?"

Courtney also saw a Chairbot on campus and thought, "'Oh, there's a person driving,' but I assumed this one wasn't that way? Possibly because I didn't see anyone [driving the shoe rack]." She also on the 3rd Monday suggests adding a remote control function, apparently without suspecting that it already has one.

When the interviewer mentioned that the wizard sometimes drove closer to the bench than normal, Bill said, "I noticed that," but he also said that he did not become suspicious that the robot was being remote controlled.

7.5.1 Why Did Not Anybody Suspect a Teleoperator? Adam thought that nobody guessed the robot was being teleoperated, because "people have this idea of what a robot can do and can't do, and the functions that you picked were well within the realm of what a robot *could* do." "If you had it do some sort of crazy functions [like] recognizing you and [saying your] name or something like that I'd be like, 'ehh. That's not real.' Or, 'that's less likely to be real...'." Frankie makes a related point in her final interview—that her mental model was limited by her understanding of technology: "I think the abilities I assume the robot has are within my limits of what I could do with a robot I could create." This is related to the earlier discussion in Section 7.3.3 about people's mental models for a "robot" as a general category.

The interviewees mentioned their existing knowledge about technology when talking about the robot. Adam and Courtney, for example, showed that they understand that computers can learn to classify things into different categories based on many training examples. Bill understood that the robot has "got a clock" that is probably "synced someplace." Adam even draws an analogy about the robot using another technology he knows about: "Yeah, so you never know, you could have a shoe rack like this that has a very innocuous job—holding on to people's shoes—but it could be keeping track of many other things. It's like license plate readers in intersections: nobody knows that they've gone through and their license plate's been scanned, but these things can scan tons of license plates at a time, and so you just never know that it's happening."

Users' understanding of what is easy, hard, and impossible with modern technology—whether they are right or wrong—could impact their judgments about whether a robot is autonomous or teleoperated. It seems likely that robots will be suspected of being teleoperated when they do things that people think are impossible to do autonomously.

Adam also described the robot's behavior as "realistic": "I think the way you had it in a very automated way that it went seemed very robotic to me." We intentionally designed the robot's behaviors to be robotic, especially at first: It drove in straight lines according to a simple, repeating pattern and typically only did one thing at a time. It will be important to understand each user's model of how robots act—i.e., of what robotlike behavior looks like—and develop ways to control how robotic a robot acts, especially if it has a large influence on whether people believe it is autonomous.

A third possible explanation is that people did not suspect a teleoperator, because they did not see one, nor did they see a wire or antenna that obviously hinted at one. This could pertain to the way humans perceive social entities: perhaps we are inclined to attribute independent autonomy to bodies that are self-contained and not connected to another social actor. By this theory, a remote-controlled car would be easier to consider autonomous than a prosthetic arm would, and a robot connected to an operator's computer by a thin wire would present a border case. Wireless, remote control would cause dissonance between our unconscious impressions and our conscious understanding. Further study to test this hypothesis could help HRI researchers understand how people estimate the degree of connection between a robot and the people who might be operating or supervising it.

7.6 Experimenting with the Robot

In general, the six interviewees did not actively experiment with the robot very much; they mostly just watched it. When they were asked a question in an interview for which they did not have an answer, they were quick to make guesses based on reasoning from what they had observed passively and from their mental models. They rarely took actions that were just to test the robot. Adam, for example, notes that although he had "opportunities," "I never tried to test out the limits of the robot, it just never really occurred to me...." In fact, Adam on the fourth Wednesday says: "I've probably never actually touched it [i.e., the body of the robot], I just touch my shoes, right?"

He contrasts this with his three-year-old daughter, who he says would be "exploring it with her hands," "she'd definitely want to sit on it," "she would probably try to pick it up," and she would "grab" and "pull on" the sensor. This exposes a limitation in our sample of interviewees: None of them were as comfortable experimenting with the robot as a young child might be. Future work could focus specifically on how young children experiment with novel robots. The cause of the lack of experimentation in our adult interviewees could also be investigated. One hypothesis is that they thought of the robot as having regions of its body that it does not want people to touch. This possibility has been previously explored in the context of humanoid robots [29], but it is unclear whether it will apply to non-anthropomorphic robots. Another possibility is that they were following what they believed to be the study team's wishes by not messing with the robot. Or, finally, they might not have thought of the robot as a social being or as the study team's property, but were instead discouraged from experimenting with it, because it might look bad in front of the other people in the hallway.

When the interviewees did experiment with the robot, some connections between their experimentation and their mental models became apparent. For example, Adam on the second Wednesday came to class late, after the robot had parked. He "tried to wave in front of where the shoe rack was," "to see if it was moving," "but it didn't move." He then interprets the results of his experiment based on his mental model of how the robot works: "Maybe I didn't hit the trigger or whatever," which also shows that his intention—to "hit the trigger"—was driven by his understanding that the robot *has* a "trigger" to hit. Hence, aspects of your mental model can inform how you create experiments as well as how you interpret the results. Presumably, someone also might design an experiment to help them decide between multiple possible models that they have brainstormed. If the result cannot easily be explained by any of the possible models, then that person might brainstorm one or more new models that do explain it. Choosing *not* to experiment can also be motivated by a mental model. Frankie in her final interview, for example, said she assumed it cannot hear, "so I didn't say anything to it."

8 ADDITIONAL QUESTIONS FOR FUTURE RESEARCH

The previous section presents interesting phenomena from the study organized into six topics—recording these phenomena was the first contribution of this work. The second contribution was to generate new, data-driven research questions to guide future work in mental model formation about robots. Many such research questions were introduced with the phenomena that inspired them in the previous section. Those have also been collected and summarized in Table 4 for easy viewing (see Section 11 below). A few additional, broader research questions are presented in this section.

What about nondeterministic behaviors? Although the wizard sometimes made mistakes, the robot's behaviors were supposed to be completely deterministic and predictable. It would be interesting to study how people reason about robots with more stochasticity (randomness) in their behaviors, or that change their rules of behavior more gradually over time (e.g., via learning). It seems from our results that people might be slow to notice that a robot's behaviors are changing, especially if they only observe it for short periods of time. Future work could explore how people make this judgment—what cues they look for, and how they reason about it.

Is there a way to present behaviors to control mental model formation? Presumably it matters which behaviors people see and which order they see them in when building a mental model of the robot. Is it possible to unveil a robot's features in a certain order to increase the accuracy of the final mental model, or help people develop mental models more quickly? For example, it might be a bad idea to show a sophisticated ability—e.g., like if your robot can speak—in the first interaction, because people might use it to infer other abilities that your robot does not have [47].

Table 4. A Summary of Research Questions for Future Study from the Discussion of the Cross-case Analysis Results in Section 7

Research question identified in cross-case analysis	Section mentioned
What factors might influence mental model improvement rate?	7.1
Might "sensemaking" theories explain participant responses to robot failure?	7.2.1
What prompts further reasoning about observed robot behaviors?	7.2.4
How useful are theories of attribution to forming mental models of robots?	7.2.4
What about a robot causes users to compare it to certain things and not others?	7.3.1
How does a participant's mental model of robots-in-general change over time?	7.3.3
How is mental model structure affected by design cues?	7.3.4
How might researchers distinguish between sincere beliefs and intentional projections?	7.3.5
What prompts users to attribute sensing abilities to robots without obvious cues?	7.4
What prompts a user to consider a robot autonomous?	7.5.1
How do children experiment with novel robots?	7.6
What causes users to avoid experimenting with the robot?	7.6
How do users decide when, and how, to experiment with robots?	7.6

Note that this is different than teaching someone about simpler behaviors first to make the more complex behaviors easier to understand—here we are talking about avoiding actions when they could have misleading *implications*, perhaps by first taking actions that eliminate those inferences as possibilities. Future work should include multi-interaction studies in which we manipulate the order in which the behaviors in the robot's repertoire are revealed. Maybe the robot would also reveal negative behaviors, too: it could intentionally fail to do something or otherwise show what is *not* in its repertoire. Understanding the impact of these manipulations would help designers shape mental model formation.

9 LIMITATIONS

A single analyst chose the interesting phenomena from the interview data and organized them into topics for presentation in Section 7. Whereas we believe each phenomenon to be presented accurately and with evidence from the transcripts, a different analyst may have chosen different phenomena as interesting enough to report, or organized these phenomena differently. This study does not benefit from the fruits of a second analysis.

This study also could have benefited from more of the eight "validation strategies" highlighted by Creswell [4]. In particular, "member checking"—i.e., getting feedback from the interviewees themselves on our preliminary interpretations of their interview responses—might have been quite useful.

Other limitations of this study have become a source of fruitful speculation by the authors, and are presented below with suggestions for future work.

Participants might pay less attention and do less thinking about the robot in a purely observational study. Several of the interviewees admitted to changing their behavior, because they knew they would be interviewed about it. If an interviewee came to their first interview and did not have much to say about the robot, then this often motivated them to spend more time watching the robot or interacting with it. For example, Courtney on the fourth Wednesday says, "It's kind of just been a progression in paying attention to it more, because I know we're doing these interviews so I pay attention to it more, I guess, instead of in passing." Elsie on the fifth Wednesday says she put her shoes on the robot when she might not have otherwise: "I noticed the shoe rack and it was pretty full, but I decided I was going to use it anyway, because I knew we were going to be talking about it....” So, at least some of the interviewees would have probably behaved differently, especially later in the study, if there were no interviews motivating them to make clear, detailed observations of the robot.

The mental models described during the interviews might not be what is driving user behavior. This analysis has focused on self-reports—i.e., on the parts of the mental model formation process of which people are consciously aware. If people understand the robot like they understand other people, however, then there is evidence that suggests they will not just do so consciously (e.g., when responding to interview questions) but also unconsciously, and that while these two processes are connected they are also distinct enough to yield different mental models [45]. To the extent that this extends to human-robot interaction, an analysis of self-report data might suggest a mental model of the robot that does not always predict the user’s behavior, because it lacks information about implicit beliefs that were formed unconsciously. In the worst case, these self-reports could merely be rationalizations with *no* predictive power.

The extent to which the findings of this study describe a mental model formation process that is predictive of user behavior is unknown. Future work in this area should make sure to collect behavioral data in addition to self-reports and compare them. For example, in this study video and audio recordings were made of all activity within the robot’s taped-off area in the hallway. Planned future work will analyze the behaviors of the six interviewees in these recordings for (in)congruence with the beliefs they expressed in the interviews.

Conversations between people about the robot could be much more important in other scenarios. People did not interact with each other much during the study. Very few of the yoga students knew each other, and most people kept to themselves while in the hallway before and after class. This means that there were not many opportunities for people to share observations or beliefs about the robot. If Donna had mentioned the possibility of someone remotely influencing the robot’s actions to Adam, for example, then he might have become much more suspicious of a remote operator than she did. It could be important to understand how these and other phenomena like rumors (which might change as they spread) and a group-constructed persona for the robot (e.g., assigning it a name and personality) could influence mental model formation about the robot.

Introducing people to a new robot could include more initial explanation or training. Participants were told almost nothing by the study team about how the mobile shoe rack looks and operates. Adam, at least, was aware of this when asked at the end of the study about how he formed his mental models of the robot: “...so all these [hypotheses about how the robot works] have come from just the observations and my interactions with it....” The study was designed so that participants would not get much additional help forming or adjusting their mental models, but in reality people often see television ads, YouTube videos, or news stories about robots before they first interact with them. They might also read written descriptions, see still photographs, or hear about the robot from a friend or colleague. This initial information about the robot could be received as authoritative (e.g., if it is from the robot’s manufacturer) or not so much. It could be important to understand how this information sets up initial expectations of the robot[1] and how people reconcile observations with rumors when the two do not match.

Limitations in demographics of interviewees. The interviewees who volunteered for this study were relatively well-educated people, several of whom worked in academia. Perhaps they were therefore more likely to hypothesize explanations for the robot’s behavior and to evaluate these explanations in light of their observations. They might also have been more likely to avoid being seen treating the robot as a person or animal if other people might see that as childish or

[1]Prior research on the effects of these expectations includes the work by Paepcke and Takayama [36].

ignorant of the fact that robots are not *really* alive. All six of the interviewees were also relatively open to new technologies according to their responses to our questionnaire items.

The choice of a university campus in a small town with low crime rates might have influenced how much people paid attention to the robot and noticed details about it. This hypothesis was suggested by Frankie, who was surprised by how many people did not "look up" at the robot or talk about it. It seems she expected them to be more alarmed by a "weird shelf moving around" and to at least ask somebody else what it was. She theorizes that people in this area "just don't notice their surroundings" compared to people from her home town, where "their surroundings aren't always perceived as safe." The extent to which people are suspicious of the robot (or its operators) or are worried about their safety or privacy around it could influence the way they form a mental model of what it is and how it works.

10 STUDY DESIGN RECOMMENDATIONS

If we were to run this study again, then we would make several key changes to the methodology.

1a. *Loosen* Experimental Control. One method of improving this study would be to loosen some experimental controls and do more active exploration. Abnormal circumstances, such as when the robot ran out of battery on the fourth Monday, provides an opportunity to see how participants react when things do not go according to plan. It would be fruitful to actually plan intentional malfunctions or failures to see how participants handle them—real robots do malfunction, after all.

It may also be useful to improvise some interactions to test the limits of a particular person's mental model of the robot. For example, if they currently believe that the robot is autonomous, we could try to find behaviors that would convince them that it is teleoperated. Or, if they believe the robot cannot hear, then the wizard could try responding to sound in different ways until they believe that it can. This would be a more active, targeted approach to understanding what influences specific elements of a mental model.

1b. Alternatively, *Tighten* Experimental Control. An alternative study design to the more exploratory one above would be to attempt a more direct comparison between participants by controlling which behaviors they experience. Perhaps the robot would wait until a participant is watching before displaying a new behavior. It might help to simplify the scenario so that only one participant encounters the robot at a time—this would also prevent the participant from seeing how other people react to the robot and how the robot interacts with them. Extra factors like whether the participant is paying attention to each behavior could also be more carefully measured so that the variance in mental model formation that is attributable to those factors could be identified during data analysis.

Additional control could also be imposed on some parts of the mental model formation process to make the study more focused. For example, participants could be manipulated to approach the robot with one particular mental model as their template. The appearance and behavior of the robot could be designed to evoke that mental model—maybe a person's existing model for a pet dog or a human butler. Pilot studies would be required to make sure that this desired impression is made clearly and uniformly across participants. Controlling this source of variance might help provide a clearer picture of other parts of the mental model formation process such as updating based on interactions with the robot. However, findings from such a study might not generalize to robots that suggest different templates or that do not suggest a template at all.

2. Avoid Implicitly Motivating Participants to Pay Attention to the Robot. It became clear that the interview process caused the interviewees to pay more attention to the robot and interact with it more. Participants often seemed uncomfortable with not having an answer to interview questions that asked for specific details about the robot, and often reasoned out an answer at that moment, as also discussed in Section 7.2.4, or said they would pay more attention to the robot next time they came to class. This was noticed in the interviewees' behaviors, but several of them also admitted it—for example, Courtney on Wednesday of Week 4 said, "I know we're doing these interviews so I pay attention to it more, I guess, instead of in passing." Also, Elsie on Wednesday of Week 6 said, "I didn't use it [the mobile shoe rack] until I had said that I was going to do the interviews...." The motivation provided by the interviewing process seems to have caused more frequent interactions with and thoughts about the robot. For this reason the findings presented in this article may not generalize to contexts in which people have low motivation or incentive to understand the robot in very much detail. For example, if the mental model formation process were more unconscious instead of deliberate when motivation is low, then there has been evidence that suggests the process might work differently than what has been reported here [45]. Future studies might focus more on contexts wherein people are not prompted or incentivized to produce high quality mental models.

In addition to just paying more attention to the robot, the interviewees also learned to focus on describing the robot's behaviors even though that was never explicitly stated to be the main focus of the study.

One way to mitigate the issue of mid-study interviews influencing participants' behavior later in the study would be to only ask interview questions at the end of the study. The mid-study interviews could be replaced with more detailed observations of participant behavior. Study scenarios and robot behaviors could be designed such that participants' reactions are easily observable and provide unambiguous information about their current mental models. Alternatively, researchers could just ask questions in the middle of the study that do not embarrass interviewees that have not been paying as much attention—e.g., "Tell me about your experience taking off your shoes today before class" might be better than "describe all of the mobile shoe rack's behaviors."

3. Use Theory from Past Literature on Mental Models. As mentioned in Section 3.1 this study was not designed according to any particular theory from the literature on mental models. In retrospect, however, we did impose our own framework to a certain extent by emphasizing (1) the robot's behavioral patterns and (2) perception capabilities during the interviews. Some responses during the interviews indicated that this might not be a good fit for how people naturally think about the robot. For example, interviewees often did not know how to answer the questions about specific capabilities like recognizing different people or sensing whether their shoes are on or off. Some of the interviewees commented that they had never thought about those questions before the interview.

To avoid imposing an ill-fitting model structure on participants, researchers could either eliminate those more assumption-laden interview questions or go the other way and design the study according to a theoretical framework that has been shown to fit in, e.g., HCI scenarios. Better yet, several theoretical approaches could be tried and compared to see which fits best when users are modelling robots. Theories about model structure in particular could help researchers create descriptions (in fact, *models*) of mental models that fit what is actually happening in participants' minds.

4. Collaborate with Design Experts. This study also revealed how influential the design of the robot's appearance and behaviors is to how participants respond to it. In future work it will become important to produce robots that target a certain phenomenon of interest by evoking a certain response in participants with high precision and consistency. For example, a robot could be designed to ambiguously suggest two different mental model templates—e.g., either a pet dog or a human butler—to study how people choose between them. Alternatively, two slightly different behavioral modes could be designed for the robot that are activated based on some feature of the environment, and see how long it takes people to connect the behaviors to that environmental feature. Future studies of mental model formation will require design experts on the research team so the robot and its interactions can be better aligned with the research goals in this way.

Longitudinal, in-the-wild studies are often avoided because of how long they take and their lack of focus. All of these recommendations are intended to make them more targeted and efficient so researchers will start doing longer, more realistic studies.

11 SUMMARY OF FINDINGS AND RECOMMENDATIONS

This article presented findings from the first long-term, in-the-wild study of mental model formation about the perceptual capabilities, rules of behavior, and communication capabilities of a novel robot. Data were analyzed from 28 interviews of six participants during six weeks of interactions with a "mobile shoe rack" working outside a yoga classroom. The findings are summarized as follows:

(1) **Different participants experienced a new robot very differently, especially over multiple interactions.** Each participant had a different set of observations: out of the robot's list of possible behaviors, only a subset were triggered while a particular participant was watching, and only a subset of these were noticed. Each participant also had a different preconception of what is technically possible for an autonomous robot. Finally, participants varied widely in motivation to figure out the robot, and improved the predictive accuracy of their mental models at very different rates.

(2) **Mental model formation was a complex process consisting of multiple components and sometimes yielding surprising results.** Several participants were noted as using an existing mental model as a "template" for their model of the mobile shoe rack. One participant knowingly maintained a make-believe persona for the robot alongside the real one. Given some observations of the robot, participants then used various types of reasoning—eight types have been described—to draw conclusions about it. Inferences were often indirect, like inferring the presence of a sensor without actually seeing it by using the robot's behavior. Participants also made predictions about the robot's behavior and designed experiments to test their mental models, although they rarely ran these experiments when given the opportunity. All interviewees were surprised to hear at the end of the study that the robot was being controlled by a human.

Additionally, a number of future research directions in mental model formation in HRI have been identified based on experiences and findings from this study. Some of these questions might be asked here in this article for the first time; all are motivated by the observed phenomena reported in Section 7. These recommended research questions are summarized here in Table 4. Additional, broader directions for future work can be found in Sections 8 and 9.

Study Design Recommendations included (1a) either improvise more with individual participants or (1b) tighten experimental control so results can be compared directly between participants; (2) avoid implicitly encouraging participants via the interview questions to pay more attention to the robot; (3) evaluate some mental model theories from the literature to see how they fit in HRI scenarios; and (4) collaborate with design experts when developing the robot's appearance and behaviors.

All of these contributions demonstrate the fruitfulness of long-term, in-the-wild studies for learning about mental model formation in HRI. Mental model formation is fundamental to many important areas of HRI research, such as user interface design, human-robot collaboration, trust, safety, and privacy. This work highlights several open questions in mental model formation and adaptation, particularly relevant to robot designers who have a vested interest in understanding how users will come to understand and use their robots. This report aims to bootstrap more research on how people form mental models of robots.

APPENDIX

A INTERVIEW GUIDE

Here we include the interview guide used in our study. See Section 5 for a description of the interview procedure and analysis.

General Prompts

- "Let's focus on the last time you came to class. When was that? Tell me what you remember about the mobile shoe rack from that day."
- (Follow-ups:)
  - "Tell me about that."
  - "Do you remember what you thought about that?" "Any other thoughts?"
  - "What did you notice?" "Notice anything else?"
  - "Did you learn anything from [that]?" "Anything else?"
  - "What stood out to you about the mobile shoe rack? What did you notice?"
- (Prompts to make sure they've covered all their interactions with the "Mobile Shoe Rack":)
  - "Did anything else happen with the mobile shoe rack that day?"
  - "Did you see any other people interact with the mobile shoe rack?"
  - "Have you had any conversations with other people about the mobile shoe rack?"
- (If this was their first experience with the shoe rack:)
  - "Do you remember what you were expecting before you first saw the shoe rack?" "How did it compare to your expectations?"

Moving beyond their experiences in the hallway...

- "What are your opinions about the mobile shoe rack? What do you think of it?"
- "What are some things you like about the mobile shoe rack? How about some things you don't like? Do you have any ideas for how to make it better?"
- "Describe to me what the mobile shoe rack does. Describe all its actions."
- "Pretend I'm a friend of yours who has never seen the mobile shoe rack and you want to describe it to me. What would you tell me?"
- "Would you want this mobile shoe rack in your home?" "Can you think of a place where the mobile shoe rack would be a really good idea?" "...how about where it would be a really *bad* idea?" "Why?"
- "Is there anything you'd like to know about the mobile shoe rack at this point? that you're wondering about? ...What else?"

If they've seen the MSR more than once:

- "You've been around the mobile shoe rack a few times now. Have you learned anything about it since the first time you saw it?"
- "Is there anything you notice now that you didn't notice at first?"

At the final meeting, right before debriefing them, we can ask about our DVs more directly:

- "Give us your best guess about the mobile shoe rack's capabilities. What can it do?"
- (ONLY if they need more prompting should you use these prompts:)
  - "Does it record/store things? Like what?"

- "Can it see? Where and how far? Hear? Where and how far? How do you know? Understand speech?"
- "Can it distinguish humans from the furniture? Like if you put a trash can in front of it?"
- "Can it tell when your shoes are on vs. off? If you're in the process of taking them off?"
- "Does it know when you've interacted with it?"
- "If you sat in front of it again, do you think it would know you already gave it your shoes?"
- "How much was it paying attention to things? Did it notice you? Other people? How much attention to detail?"
- "Do you think it remembered you from day to day?"
- "Whether you're late? Can it tell how you look in your workout clothes? Whether you were awkward or clumsy or slow? If not, did you feel awkward anyway?"
- "How did you figure that out? How long did it take you? Tell us the story."

ACKNOWLEDGMENTS

The study team wants to thank Dr. Sam Logan's lab for their welcome and the use of their space. We also thank Sonny Goodnature for helping us set up our equipment. Thanks also to "Bill" and "Frankie" for opening their classroom, and to all six interviewees—including "Adam," "Courtney," "Donna," and "Elsie"—for their time, openness, and thoughtfulness. The first author wants to thank his colleagues for helping to design this study: Duy Nguyen, Sabrina Bradshaw, Alison Shutterly, Wendy Xu, and Marlena Fraune.

REFERENCES

[1] Christine L. Borgman. 1999. The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *Int. J. Hum.-Comput. Studies* 51, 2 (Aug. 1999), 435–452.
[2] J. M. Carroll and J. C. Thomas. 1982. Metaphor and the cognitive representation of computing systems. *IEEE Trans. Syst. Man Cybernet.* 12, 2 (Mar. 1982), 107–116.
[3] Elizabeth Cha, Anca D. Dragan, and Siddhartha S. Srinivasa. 2015. Perceived robot capability. In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN'15).* IEEE, 541–548.
[4] John W. Creswell and Cheryl N. Poth. 2016. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches.* SAGE Publications, New York, NY.
[5] Maartje De Graaf, Somaya Ben Allouch, and Jan Van Dijk. 2017. Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction.* ACM, 224–233.
[6] D. C. Dennett. 1971. Intentional systems. *J. Philos.* 68, 4 (1971), 87–106.
[7] D. C. Dennett. 1989. *The Intentional Stance.* MIT Press.
[8] Ylva Fernaeus, Maria Håkansson, Mattias Jacobsson, and Sara Ljungblad. 2010. How do you play with a robotic toy animal?: A long-term study of Pleo. In *Proceedings of the 9th International Conference on Interaction Design and Children.* ACM, 39–48.
[9] Julia Fink, Valérie Bauwens, Frédéric Kaplan, and Pierre Dillenbourg. 2013. Living with a vacuum cleaning robot. *Int. J. Soc. Rob.* 5, 3 (2013), 389–408.
[10] Robert S. Fish, Robert E. Kraut, Robert W. Root, and Ronald E. Rice. 1992. Evaluating video as a technology for informal communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 37–48.
[11] David C. Funder. 1995. On the accuracy of personality judgment: A realistic approach. *Psychol. Rev.* 102, 4 (1995), 652.
[12] Susan R. Fussell, Sara Kiesler, Leslie D. Setlock, and Victoria Yew. 2008. How people anthropomorphize robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI'08).* IEEE, 145–152.
[13] Michael J. Gill, William B. Swann Jr., and David H. Silvera. 1998. On the genesis of confidence. *J. Personal. Soc. Psychol.* 75, 5 (1998), 1101.
[14] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C. Schultz, et al. 2005. Designing robots for long-term social interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05).* IEEE, 1338–1343.

[15] Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Affective personalization of a social robot tutor for children's second language skills. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI'16).* 3951–3957.

[16] Frank Halasz and Thomas P. Moran. 1982. Analogy considered harmful. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'82).* ACM, New York, NY, 383–386.

[17] Edward E. Jones. 1990. *Interpersonal Perception.* WH Freeman/Times Books/Henry Holt & Co.

[18] David E. Kieras and Susan Bovair. 1984. The role of a mental model in learning to operate a device. *Cogn. Sci.* 8, 3 (July 1984), 255–273.

[19] Sara Kiesler. 2005. Fostering common ground in human-robot interaction. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'05).* IEEE, 729–734.

[20] Sara Kiesler and Jennifer Goetz. 2002. Mental models of robotic assistants. In *Proceedings of the Extended Abstracts on Human Factors in Computing Systems (CHI'02).* ACM, 576–577.

[21] Kheng Lee Koay, Dag Sverre Syrdal, Michael L. Walters, and Kerstin Dautenhahn. 2007. Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN'07).* IEEE, 564–569.

[22] Andreas Kolling, Phillip Walker, Nilanjan Chakraborty, Katia Sycara, and Michael Lewis. 2016. Human interaction with robot swarms: A survey. *IEEE Trans. Hum.-Mach. Syst.* 46, 1 (Feb. 2016), 9–26.

[23] Minae Kwon, Malte F. Jung, and Ross A. Knepper. 2016. Human expectations of social robots. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI'16).* IEEE, 463–464.

[24] Min Kyung Lee, Karen P. Tang, Jodi Forlizzi, and Sara Kiesler. 2011. Understanding users' perception of privacy in human-robot interaction. In *Proceedings of the 6th International Conference on Human-robot Interaction.* ACM, 181–182.

[25] Sau-lai Lee, Ivy Yee-man Lau, Sara Kiesler, and Chi-Yue Chiu. 2005. Human mental models of humanoid robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'05).* IEEE, 2767–2772.

[26] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social robots for long-term interaction: A survey. *Int. J. Soc. Robot.* 5, 2 (2013), 291–308.

[27] Séverin Lemaignan, Julia Fink, Pierre Dillenbourg, and Claire Braboszcz. 2014. The cognitive correlates of anthropomorphism. In *Proceedings of the Human-Robot Interaction Conference Workshop on HRI: A Bridge Between Robotics and Neuroscience.*

[28] Daniel T. Levin, Caroline Harriott, Natalie A. Paul, Tao Zhang, and Julie A. Adams. 2013. Cognitive dissonance as a measure of reactions to human-robot interaction. *J. Hum.-Robot Interact.* 2, 3 (2013), 3–17.

[29] Jamy Jue Li, Wendy Ju, and Byron Reeves. 2017. Touching a mechanical body: Tactile contact with body parts of a humanoid robot is physiologically arousing. *J. Hum.-Robot Interact.* 6, 3 (Dec. 2017), 118–130.

[30] Joseph A. Maxwell. 2012. *Qualitative Research Design: An Interactive Approach.* Vol. 41. SAGE Publications, New York, NY.

[31] Brian Ka-Jun Mok, Stephen Yang, David Sirkin, and Wendy Ju. 2015. A place for every tool and every tool in its place: Performing collaborative tasks with interactive robotic drawers. In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN'15).* IEEE, 700–706.

[32] Neville Moray. 1999. Mental models in theory and practice. *Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of Theory and Application* (1999), 223–258.

[33] Richard E. Nisbett and Timothy D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychol. Rev.* 84, 3 (1977), 231.

[34] Don Norman. 2013. *The Design of Everyday Things: Revised and Expanded Edition.* Constellation.

[35] Donald A. Norman. 2014. Some observations on mental models. In *Mental Models.* Psychology Press, 15–22.

[36] Steffi Paepcke and Leila Takayama. 2010. Judging a bot by its cover: An experiment on expectation setting for personal robots. In *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction.* IEEE Press, 45–52.

[37] Stephen J. Payne. 2003. Users' mental models: The very ideas. In *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, John M. Carroll (Ed.). Morgan Kaufmann, Burlington, MA, 135–156.

[38] Elizabeth Phillips, Scott Ososky, Janna Grove, and Florian Jentsch. 2011. From tools to teammates: Toward the development of appropriate mental models for intelligent robots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55. SAGE Publications, Los Angeles, CA, 1491–1495.

[39] Aaron Powers and Sara Kiesler. 2006. The advisor robot: Tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction (HRI'06).* ACM, New York, NY, 218–225.

[40] A. Powers, A. D. I. Kramer, S. Lim, J. Kuo, Lee Sau-lai, and S. Kiesler. 2005. Eliciting information from people with a gendered humanoid robot. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'05).*158–163.

[41] Selma Sabanovic, Marek P. Michalowski, and Reid Simmons. 2006. Robots in the wild: Observing human-robot social interaction outside the lab. In *Proceedings of the 9th IEEE International Workshop on Advanced Motion Control*. IEEE, 596–601.

[42] Burkhard Schafer and Lilian Edwards. 2017. "I spy, with my little sensor": Fair data handling practices for robots between privacy, copyright and security. *Connect. Sci.* 29, 3 (2017), 200–209.

[43] Mark Scheeff, John Pinto, Kris Rahardja, Scott Snibbe, and Robert Tow. 2002. Experiences with sparky, a social robot. In *Socially Intelligent Agents*. Springer, Berlin, 173–180.

[44] Matthias Scheutz, Scott A. DeLoach, and Julie A. Adams. 2017. A framework for developing and using shared mental models in human-agent teams. *J. Cogn. Eng. Decis. Mak.* 11, 3 (2017), 203–224.

[45] Dana Schneider, Virginia P. Slaughter, and Paul E. Dux. 2015. What do we know about implicit false-belief tracking? *Psycho. Bull. Rev.* 22, 1 (2015), 1–12.

[46] David Sirkin, Brian Mok, Stephen Yang, and Wendy Ju. 2015. Mechanical ottoman: How robotic furniture offers and withdraws support. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 11–18.

[47] William D. Smart and Neil M. Richards. 2014. How the law will think about robots (and why you should care). In *Proceedings of the IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO'14)*. IEEE, 50–55.

[48] Marco Spadafora, Victor Chahuneau, Nikolas Martelaro, David Sirkin, and Wendy Ju. 2016. Designing the behavior of interactive objects. In *Proceedings of the 10th International Conference on Tangible, Embedded, and Embodied Interaction (TEI'16)*. ACM, 70–77.

[49] Kristen Stubbs, Debra Bernstein, Kevin Crowley, and Illah R. Nourbakhsh. 2005. Long-term human-robot interaction: The personal exploration rover and museum docents. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED'05)*. 621–628.

[50] JaYoung Sung, Henrik I. Christensen, and Rebecca E. Grinter. 2009. Robots in the wild: Understanding long-term use. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*. ACM, 45–52.

[51] Karl E. Weick. 1995. *Sensemaking in Organizations*. Vol. 3. SAGE Publications, New York, NY.

[52] Robert H. Wortham, Andreas Theodorou, and Joanna J. Bryson. 2016. What does the robot think? Transparency as a fundamental design requirement for intelligent systems. In *IJCAI-2016 Ethics for Artificial Intelligence Workshop*.

[53] Stephen Yang, Brian Ka-Jun Mok, David Sirkin, Hillary Page Ive, Rohan Maheshwari, Kerstin Fischer, and Wendy Ju. 2015. Experiences developing socially acceptable interactions for a robotic trash barrel. In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN'15)*. IEEE, 277–284.

[54] Robert K. Yin. 2017. *Case Study Research and Applications: Design and Methods*. SAGE Publications, New York, NY.

[55] Tom Ziemke, Serge Thill, and David Vernon. 2015. Embodiment is a double-edged sword in human-robot interaction: Ascribed vs. intrinsic intentionality. In *Proceedings of the Workshop on Cognition: A Bridge Between Robotics and Interaction at the ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. 1–2.

[56] Jakub Złotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. 2015. Anthropomorphism: Opportunities and challenges in human–robot interaction. *Int. J. Soc. Robot.* 7, 3 (2015), 347–360.

Received February 2019; revised September 2020; accepted November 2020