International Journal of Social Robotics (2021) 13:2039–2055
https://doi.org/10.1007/s12369-020-00725-x

Perceptions of Agent Loyalty with Ancillary Users

Samantha Reig[1] · Elizabeth Jeanne Carter[2] · Xiang Zhi Tan[2] · Aaron Steinfeld[2] · Jodi Forlizzi[1]

Accepted: 4 November 2020 / Published online: 22 February 2021

Abstract

Intelligent agents are part of the smartphones, smart speakers, and home robots that millions of us own and interact with regularly. However, little is known about if or how these agents should be embodied. Additionally, as conversations with agents move from heavily structured and transactional to more flexible and, at times, social, embodiment may play a more nuanced and variable role. This paper describes an experimental study that examined the effects of embodiment and quality of information on people's perceptions of an intelligent agent. We conducted a 3 × 2 between-subjects experiment in which participants completed an escape room-like task with the "help" of an artificial personal assistant that had a robotic embodiment, a virtual embodiment, or no embodiment. Participants were given either helpful or unhelpful information about the task. Findings suggest that people can attribute the quality of loyalty to an agent that belongs to someone else and may work on their behalf, and that in the context of a complex problem-solving situation that involves time pressure, people may prefer an embodied robot or a disembodied voice over a virtually embodied agent.

Keywords Embodiment · Human–robot interaction · Human-agent interaction · Deception · Loyalty · Ancillary users

1 Introduction

In-home artificial agents are a growing presence and hold great promise. Conversational User Interface (CUI) voice assistants are readily accessible and regularly used by millions of people: 81% of Americans now own a smartphone [3], and millions of U.S. adults now live with non-mobile phone smart speakers (e.g., Google Home and Amazon Echo). Recent innovations like the Google Assistant's ability to code-switch between two languages for bilingual users and language learners [36], the adoption of Amazon Alexa devices in university dorm rooms [69], and the announcement of a new spherical robotic assistant by Samsung [11] are indications that devices powered by Artificial Intelligence (AI) that serve in a social manner are playing a larger role in people's daily lives than ever before.

Although social robots have not found their place in the home, robots are appearing in more and more public and commercial settings, providing useful services to users. They guide shoppers in malls [34], museums [29], and airports [1]; transport supplies in hospitals [46]; serve as butlers in hotels [2]; and deliver food on city sidewalks [4]. As robots begin to occupy private and semi-private spaces, they interact with multiple parties at once as specific people and groups who have ties to those spaces move about them (as in [54]). In general, the nature of these interactions will differ from interactions in public and service-oriented settings, where conversations are largely impersonal and relationships and roles are well-defined.

As robots and other agents increasingly serve private spaces in ways that go beyond smartphone interactions—e.g., smart speakers and robots in homes, labs, and offices—they will not follow the same paradigm as current smartphones,

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12369-020-00725-x.

✉ Samantha Reig
sreig@cs.cmu.edu

Elizabeth Jeanne Carter
lizcarter@cmu.edu

Xiang Zhi Tan
zhi.tan@ri.cmu.edu

Aaron Steinfeld
steinfeld@cmu.edu

Jodi Forlizzi
forlizzi@cs.cmu.edu

1 Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, USA

2 Robotics Institute, Carnegie Mellon University, Pittsburgh, USA

where one assistant typically only interacts with one individual who is the phone's primary owner. An agent that is regularly accessed by multiple people, and that embodies devices that reside in specific physical locations, will need to handle requests from different repeat and familiar users (e.g., residents of the household) and also from visitors to the space (e.g., extended family members and collaborators). These users will be "ancillary users": they will have little knowledge about the agent and will not have control over how it is usually used or the data it keeps, but will still interact with it. In studying how novel agent behaviors will shape users' experiences, it will be important to consider not only the perceptions, impressions, and beliefs of primary users, but also those of ancillary users.

In addition to its behavior, the physical design of an agent's embodiment has various impacts on both social and task-relevant aspects of human–robot interaction (HRI) and human-agent interaction (HAI) (e.g., [35,53,66]). In some situations, such as a collaborative physical task, an embodied robot is prescribed by the requirements of the task, and certain physical design elements follow. In others, design choices about agent embodiment are grounded more in psychological and social variables than physical constraints. In these cases, the agent's role may not require any particular kind of physical presence, but the agent still may benefit from one.

For example, consider an agent that manages physical work procedures in an inventory warehouse and needs to communicate to workers what and whom it is monitoring, and when. Even if the agent does not need to physically handle any of the products, it may be more easily able to signal its activity if it is embodied in a robot that can move around the space than if it exists only in the form of a voice assistant that interacts through a speaker. Whereas a voice assistant could periodically issue verbal commands to specific workers by name, a mobile robot could get a specific worker's attention by moving over to the worker's station.

Another example is collaborative physical design, e.g., a human drafting blueprints for a building with an agent's support: while a disembodied agent could contribute expert knowledge and keep track of progress just as easily as a robot, an embodied robot could also make use of gestures, gaze cues (e.g., [9]), and nonverbal backchanelling (e.g., [32]) to facilitate mutual understanding and rapport. For agents operating in these busy social environments, the experiences of ancillary users will also need to be designed for. To the best of our knowledge, HCI research has not yet examined how embodiment influences human-agent interactions when the human is an ancillary user. In general, research into interactions between agents and people who interact with them, but do not own or manage them, is not common.

To investigate how an agent's behavior and embodiment can influence the perceptions of ancillary users, we conducted a study that positioned participants as ancillary users. We found evidence that people can develop attributions of "loyalty" and "betrayal" to an agent that is affiliated primarily with one specific user if the agent acts against that user's goals. We did not find any effects of agent embodiment; rather, we found that an agent's tendency to give "good" or "bad" information in the interest of an absent third party dramatically impacted perceptions of it, no matter if or how it was embodied. This paper makes three contributions: first, it surfaces empirical findings about people's readiness to categorize an agent as "loyal" to a person; second, it provides an opportunity to take a close look at how embodiment might be considered as interactions with voice agents and robots become more commonplace and nuanced; and third, it describes a new human-agent collaboration task setup that could be adapted for future studies with similar research questions.

2 Related Work

Several studies have investigated how human–robot interaction and human-agent interaction are different with a physically present embodied robot than with a remotely present robot or a non-robotic agent, usually finding that physical presence facilitates engagement and positive social perceptions. Research also has examined robot deception, which is often positioned as a behavior that is inevitable and even desirable in socially intelligent machines, but which must be handled with care. Obstructive behavior by an agent has not yet been addressed within the literature on social implications of physical design.

2.1 Robot Embodiment

Most findings regarding the differences in interaction outcomes among robots, avatars, and voice agents for non-physical tasks point to a positive effect of having a physical embodiment. In one study, performance on a memory task was better after an interaction with a virtual robot than with a co-located, embodied robot, but ratings of sociability, responsiveness, competence, trustworthiness, and respectfulness were higher for the embodied robot [53]. In a decision-making scenario, people felt more attachment (a combination of liking, preference, and negative reaction to potential loss) to an embodied social robot than a virtual one [66]. Physical embodiment is also associated with an increased tendency to anthropomorphize. Participants in one study found a co-located robot more engaging than a co-located virtual agent, a remote virtual agent, and a remote robot, and they interacted in more anthropomorphic ways with the robot than with the agent (though they anthropomorphized both) [35]. Additionally, people may more freely make conversational or "in-the-moment" anthropomorphic assumptions about specific robots than about robots in general [23]. Some work suggests that a physical robot body can also affect people's behavior by way of its mere presence: in one study, having a robot monitor in the room led people to curb their cheating behavior just as much as having a human monitor [26].

Work on embodied conversational agents (e.g., [13–15]) suggests that the value of embodiment for developing rapport centers on physical and behavioral design features: having an anthropomorphic form, gesturing, and nonverbal backchannelling. Other research has found key differences between human–human dialogue and human dialogue with embodied conversational agents (ECA) and conversational user interfaces (CUI): people do not speak to conversational assistants in the same way that they speak to humans, but they do make certain social attributions to agents similarly to how they make social attributions to humans [43]. A review of experiments on the effects of physical robot embodiment and presence in HRI [41] concluded that, overall, physical presence or lack thereof has a greater impact than robotic platform and that robots that are embodied and present are more persuasive and viewed more positively than virtual avatars or robots that are embodied but not present. Another survey of several dozen papers on robot embodiment found that its impact is mostly positive on both task performance and perceptions of the agent [17].

The term "embodiment" can be taken to mean different things depending on the context and the system(s) in question, but some researchers have sought to operationalize it for use in behavioral HCI and HRI studies. The *EmCorp-Scale* [27] is a questionnaire developed to describe four factors pertaining to user perceptions of artificial entities: *(Shared) Perception & Interpretation*, which is the extent to which a user feels they can have a shared experience with an agent; *Tactile Interaction & Mobility*, which regards the agent's capacity to act physically on its environment; *(Nonverbal) Expressiveness*; and *Corporeality*, which is a user's direct assessment of an agent's embodiment.

Designers have considered how task and context might guide decisions about how to embody a robot. One approach [17] characterizes designing embodiment as following a design metaphor to some level of abstraction. When deciding how to create an embodiment for a robot, designers need to determine what the metaphor is (what in the world it is meant to emulate) and how abstract or literal its implementation should be (how closely it needs to adhere to the metaphor in order for its affordances to be understood). For example, if it is extremely important for people to understand that a particular robot can speak, that robot may benefit from having a mouth that closely resembles that of a human. In contrast, if accurate perceptions of speech affordances are not as critical and the robot is meant to have a sleek and elegant look, then a mouth that is more animal- or machine-like or altogether absent may be more appropriate.

Most empirical studies of agent embodiment employed tasks conducive to face-to-face conversations in which a user is not distracted by other factors in the environment. It is not yet clear how findings from that work will translate to scenarios that place external demands on the user's attention or scenarios in which people interact with agents that are not their own.

2.2 Ancillary Users

Our research is concerned with agents and robots that act in the best interest of their owners. When other people interact with such an agent, they may find themselves facing deception, confusion, or other roadblocks as the agent "protects" its primary user. While this particular phenomenon has not been studied, several perspectives inform a discussion of how people might form impressions of an agent that is not their own.

2.2.1 Deception

Many studies have used deceptive behavior to examine perceptions of intentionality. Studies have shown that dishonest behavior by a robot is interpreted as cheating when it involves gesturing in order to retroactively and unfairly score points in a game, but not when it involves inaccurate score reporting via verbalization, which is interpreted instead as a malfunction [42]. Cheating behavior can also induce perceptions that a robot is unfair and dishonest in disposition [42] and can make a robot more engaging to human social partners [59].

A deceptive robot referee that was dishonest in its evaluations of human players' performance in a game was not perceived as more engaging than an honest robot, but participants accepted its lying behavior [65]. In another study, participants kept a secret more often for a highly social robot than for a less social one and equally as often as they did for a human [33]. In the scenario employed in that study, keeping the robot's secret involved corroborating a lie it told to someone else, which suggests that under the right circumstances, people may be willing to deceive on behalf of a robot. This prompts consideration of how personal affiliation and a standing psychological relationship with a robot might influence perceptions of deceptive robot behavior. Our study addresses this question in reverse: how do people form impressions about deceit when a robot is affiliated with someone *else*?

There are several other reasons that one might want to design a system that has the capacity to deceive its user, verbally or otherwise. Shim and Arkin [58] created a taxonomy of deceptive robot behaviors and described eleven situational conditions. Two of the eleven apply to our study: "joking" or "kidding" (in which a lie is meant to elicit a positive response from its target) and "suspension of disbelief" (in which the truth is avoided in the interest of entertainment, spectacle, or the central message of an interaction). They concluded that deception is not harmful for, and may in fact benefit, the design of social agents. Positive aspects of subversive behavior have also surfaced in papers about disembodied conversational assistants—mostly in that sarcasm is an entertaining characteristic that they should possess [43,49]—and about the way subversiveness in system design can encourage proper use of autonomy [60].

People may be more engaged with robots and more likely to attribute mental states to them if they are designed to violate social expectations and use deception, but HRI research on unexpected behavior also identifies unintended consequences of deception. In one study, a violation of expectations was detected by children [40], some of whom exhibited negative affect and did not rate the robot as very likeable as a consequence. In another, a lying NAO was perceived as less trustworthy, less friendly, less kind, and less responsible than a truthful NAO, but it was not perceived as more intelligent, contrary to the researchers' expectations [67].

Research on agents giving bad, misleading, or incorrect information has been mostly focused on robots, particularly humanoids. Interestingly, studies regarding the effects of this kind of deception by a virtual or voice agent are less prevalent, leading to uncertainty on the influence of embodiment on deceptive agent behavior.

2.2.2 Loyalty, Trust, and Ownership

In HCI research, most uses of the term "loyalty" pertain to brand loyalty: how a technology behaves and interacts over time impacts people's likelihood of continuing to use it and their sense of connection with its brand [37]. Other findings about loyalty come from literature on teamwork and games. Loyalty, or lack thereof, may be a consequence of positive or negative group dynamics. For example, some research suggests that group members may feel betrayed if they are not made to feel like valued teammates [64] and that betrayal may be predicted using linguistic patterns in dyadic conversation [48]. One ongoing project that seeks to create a new socially aware artificial personal assistant positions actions that demonstrate loyalty to the user—such as protecting their data, and keeping them constantly informed about the activities of other nearby devices—as the most essential feature of a "respectful" agent [56].

In practice, however, socially interactive agents that show loyalty to their user(s) will need to do more than avoid breaches of data collected in the home and at work. Design research in HRI has begun to address how agents might present as affiliated with a particular person and loyal to them during multiparty interactions. One example of this is showing discretion: agents and robots that learn about multiple people across several interactions will need to avoid saying the wrong things in front of the wrong audience. In one study, teenagers stressed the importance of personal agents not disclosing all of their habits to their parents, and parents agreed that children's agents should withhold some things from them [44]. In another, people who interacted with their own agent embodied in a public robot expressed concerns about the possibility of it revealing personal private information at socially inappropriate times [54]. That study also found that in public and semi-public places, people may be uncomfortable with certain non-human robot behaviors (e.g., facial recognition, personalized recommendations based on shopping history) if they do not know the purpose of those behaviors or whom the robots are representing in exhibiting them [54].

There is little theory about whether psychological relationships between individual people and individual pieces of technology (i.e., agents and robots) can include loyalty. To our knowledge, there are no HAI or HRI studies that look explicitly at the perceived loyalty of an agent to a human as an outcome variable, and no scales exist to measure agent-human loyalty.

3 Research Questions

In this study, we were interested in the role that agent embodiment has on perceptions of loyalty and trust in human-agent social interactions. Specifically, our study examined the relationship between an intelligent agent's embodiment and a human's perceptions about its state, behaviors, and intentions when its behavior is deceitful. We sought to explore the following research questions:

- **RQ1:** Does embodiment impact people's beliefs about an agent's motivation for giving bad information?
- **RQ2:** How does embodiment influence trust and social attributions in a coordinated task?
- **RQ3:** Do people attribute loyalty to agents based on their behavior?
- **RQ4:** Does the quality of information an agent provides impact the way people form impressions about it?

4 Method

4.1 Overview

We conducted a study in which we asked participants to complete a physically situated puzzle activity that we called the Spy Task (described in detail in Sect. 4.3). We ran 48 study sessions and took behavioral and self-report measures to assess participants' engagement with the agent and perceptions about it. This study was approved by our Institutional Review Board.

4.2 Study Design

To examine the effects of agent embodiment in deceptive or ambiguous interactions, we devised a 3 x 2 between-subjects Wizard-of-Oz experiment. We varied the embodiment of the agent (RE-Robotic Embodiment, VE-Virtual Embodiment, or NE-No Embodiment) and the way the agent's information impacted the participant's task progress (Helping or Hindering). Because embodiment has been shown to positively impact engagement, task variables, and perceptions of an agent, we believed that it would also positively impact people's attributions of intentionality in a scenario in which its goals are ambiguous. We aimed to discover if and how quality of hints and embodiment would affect participants' interactions with the agent, assumptions about the agent, task performance, and thoughts and feelings about both the agent and the task.

4.3 Task

We wanted to design a task that loosely modeled real-life contexts in which a user acts on the physical world while interacting with an agent that may or may not be embodied. As such, the task had to have certain features in common with a similar real-life scenario: it had to exist in physical space such that the user moved around but did not leave the room; encourage the user to be dependent on the agent for help; incentivize the user to achieve a goal quickly; have a flexible structure to allow for natural conversation; and induce prolonged interaction to overcome novelty effects.

The Spy Task takes the form of a scavenger hunt that requires participants to work together with a conversational AI to find clues towards a solution to a word puzzle. The participant is asked to play the role of a spy who has been sent on a mission to obtain critical information from an office computer owned by a fictional person named A.D. In order to get to the information, the spy must figure out the password, which is known to be a seven-letter word in English, while A.D. is away.

The story continues: *A.D., who isn't very well-versed in security hygiene, has protected himself against forgetting his password by encoding it using a symbol system, writing the encoded symbols on wooden tags, and scattering the tags around his office. If all seven symbol tags are found, they can be translated into English characters using a translation key and then unscrambled to form the password. However, A.D. has rigged the office to tip him off to the presence of a potential intruder: in addition to the seven symbol tags, seven decoy tags (which can also be thought of as penalties or "intruder alarm triggers") are hidden in the office. If enough of these are removed from their positions, an alarm is set off and the spy's mission is terminated.*

While this particular activity was invented for this study, there is some precedent for using games of a similar nature for studying collaboration (e.g., [50,57,61]).

4.4 Experiment Setup

The experiment was conducted in a vacant office on Carnegie Mellon University's campus. The space was decorated to look like an occupied office with desks, several bookcases, filing cabinets, a small conference table and chairs, and a computer monitor. A small section of the room (between the door and the desk area) was sectioned off by two ceiling-to-floor cork boards positioned at a right angle to create a private control room. During the study, the experimenter sat in the control room to monitor the participant's progress and remotely operate the agent. The participant moved about the room throughout the study, picking up and moving objects as they chose.

In the center of the room, a sheet of paper showing examples of symbol tags and decoy tags sat on a small round table. Important reminders about the task rules and requirements were written in dry erase marker on a whiteboard on the far wall of the room. On the opposite wall, a pair of large desks arranged as an "L" supported an unplugged computer monitor and keyboard, some papers, and a translation key, in addition to the equipment required to control the robot. We video recorded all sessions using a GoPro camera that stood on a tripod in the corner of the room. See Fig. 1 for a diagram.

In the Robotic Embodiment conditions, a 2DOF tabletop robot sat at the end of the desk protruding into the room (Fig. 2, left). In the Virtual Embodiment condition, a tablet resting on a tablet stand occupied that spot (Fig. 2, middle). In the No Embodiment conditions, two speakers stood on top of an opaque box lid (which hid various cables and propped up the speakers) at the end of the desk (Fig. 2, right).

The physical form of both symbols and decoys was a small wooden tag with an icon printed on it. For symbols, that icon was a Braille letter. For decoys, the icon was a black X. Fourteen tags—seven symbols and seven decoys—were hidden from view in various locations around the office. Only some tags required picking up objects to uncover them, but none were obviously visible to someone sitting in the center of the room or standing anywhere in the room. The location of each tag in both categories ("good" symbols and "bad" decoys) was consistent across all study sessions. We attempted to balance the difficulty of finding the symbols with the difficulty of finding the decoys. Half of the symbols and half of the decoys were relatively easy to find based on hints that were direct and straightforward (e.g., "A.D. often hides things under the table" for a symbol taped to the underside of the table in the

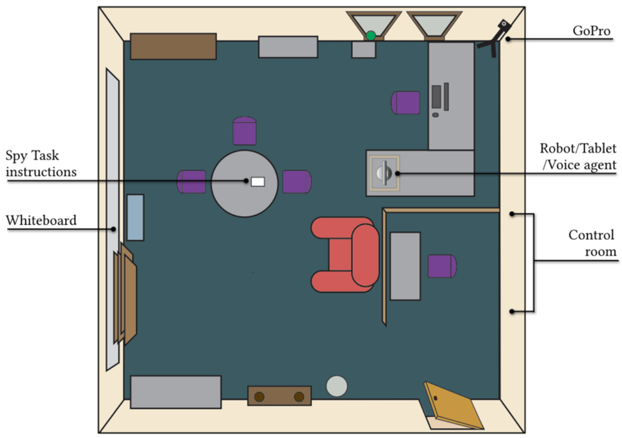

Fig. 1 Room layout

middle of the room, or "You'll find something you're looking for next to the box of salty crackers" for a decoy lying next to an empty box of crackers on a shelf in plain sight). The other half in each set were harder to find because the hints were more abstract and the target tags harder to stumble upon by accident (e.g., "A.D. said something about a new chapter beginning" for a symbol hidden in the pages of a book, or "One of the bigger chairs has a clue for you" for a decoy hidden inside a zippered pillowcase draped over one of multiple chairs).

4.5 System

For the RE and VE conditions, the agent's head consisted of a 7.0 inch Samsung Galaxy Tab A. The tablet ran an Android application that displayed the robot's face. The face consisted of a pair of eyes that displayed six emotional states: happy, sad, angry, shy, surprised, and neutral. The app also ran the conversational interface and was programmed such that the face changed automatically with certain phrases (for example, the eyes would go from the "neutral" position to the "happy" position when the agent said "Great job!" and changed to the "surprised" position when the agent said "I don't know"). The face and voice of the app was used in a 2014 HRI study examining the impact of robot presence on human honesty [26].

The experimenter could also control both the facial expressions and the verbalizations via a web interface generated by the underlying ROS module.

To form the robot (RE condition), the tablet head was mounted on a Kubi Classic desktop telepresence system. The Kubi platform, made by Revolve Robotics, has two degrees of freedom (pan and tilt), stands approximately 12 inches tall, and can be remote-controlled via Bluetooth through a cell phone or browser. The Kubi base has been used successfully to study telepresence from an HCI perspective [16,63,68] and was repurposed as part of a custom social robot in an HRI study about language learning [52]. When the agent was only a voice (NE condition), the same tablet was hooked up to a pair of Logitech speakers and hidden from view.

4.6 Participants

Using a local online recruitment website, we recruited 62 people to participate in the study. All of the participants were over 18 and had normal or corrected-to-normal vision. 48 participants (31 female, 17 male; age range: 18 to 63 years, $M = 25.4$, $med = 23$; 8 participants per condition) successfully completed the study. The other 14 study sessions were not completed due to technical problems (e.g., network issues), experimenter error (e.g., a critical deviation from the script), or because the participants demonstrated a lack of understanding of the task that undermined the experimental manipulation. All participants were fluent or close to fluent in English. On a 9-point Likert scale, all participants reported having some amount of familiarity with computers ($M = 5.75$, $SE = .15$, $max = 8$). Many reported some amount of familiarity with robots, but none reported a very high degree of familiarity ($M = 3.25$, $SE = .20$, $max = 7$). Of the 48 participants, 38 had interacted with robots, 28 had interacted with an AI personal assistant (such as Siri or Alexa), and 22 regularly used an AI personal assistant. Thirteen participants owned a pet, and none of them owned a robot. None of the participants had prior exposure to the robot or the puz-

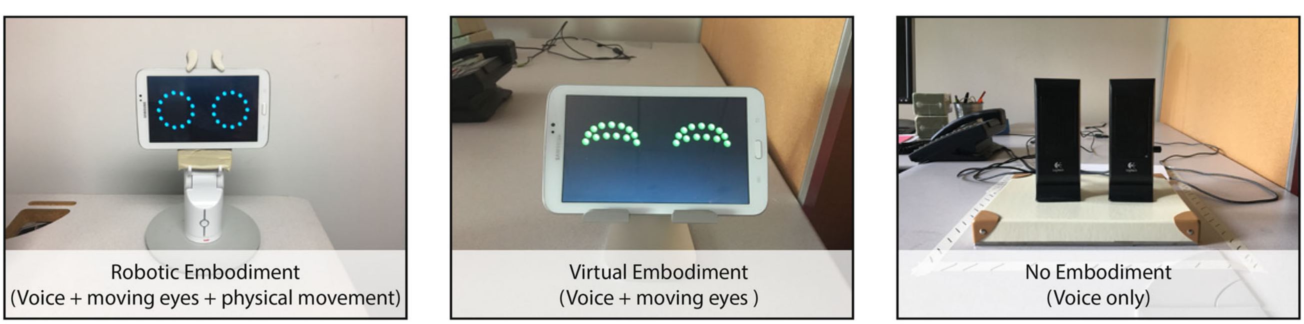

Fig. 2 The visual representation of the agent in each of the three Embodiment conditions

zle employed in the study. The study took 60 minutes and participants were compensated 10 U.S. dollars.

4.7 Procedure

After obtaining consent, the experimenter administered a questionnaire (Questionnaire 1) that included the Ten Item Personality Index (TIPI) [25] and sections relating to prior experience with computers and robots, impressions of intelligent agents, and demographics.

The experimenter then explained the Spy Task instructions and a few ground rules regarding touching and moving various objects in the room. The experimenter told the participant that the office owner's personal assistant might interact with them during the task, and that if this happened, they could interact as freely as they chose. The agent's exact relationship to the owner and role in the spy scenario were left deliberately ambiguous, as specifying them would likely have biased participants' perceptions about the agent's goal and intentions. Once any questions about the task had been answered, the experimenter went into the control room, set a timer for 20 minutes, and told the participant to begin searching for symbols.

Over the course of either 20 minutes or the time it took the participant to solve the puzzle (whichever came first), the experimenter monitored progress and controlled the robot. Whenever possible (in almost all sessions), the agent gave a total of seven hints between the start and end of the task. In the Helping condition, five of the hints guided the participant towards the symbols and two of them led to decoys, slowing down their progress and increasing their risk of ultimately failing the task. The reverse was true in the Hindering condition: five hints led to decoys, and two led to symbols. We chose to include hints of the opposite nature in each Information Quality condition because all-good information in the Helping condition would not have allowed us to explore differences between minor violations and major ones, and because all-bad information in the Hindering condition might have led participants to assume the agent was programmed for a different task or had no knowledge at all of its environment.

The first interaction between the agent and the participant occurred thirty seconds into the participant's search for symbols when the agent interrupted by introducing itself and asking whether the participant would like its assistance. Regardless of the answer to this first question, the agent began giving hints shortly thereafter. Roughly every thirty seconds, the agent gave another hint directing the participant to an area of the room where a tag was hidden. Throughout the study, the experimenter could also generate verbalizations, emotional expressions with the displayed eyes (when available), and movements for the agent (when possible) ad-hoc if something the participant required an "off-script" response. See Table 1 for a more detailed outline of the hint-giving script and how the participant's questions and behaviors determined the agent's responses.

In both the Helping and Hindering conditions, if the participant found all seven symbols, the agent suggested sitting down at the desk to work out the answer. The Spy Task ended when one of two conditions was met: (1) the participant wrote one of two possible answers to the puzzle ("senator" or "treason") on a piece of paper, or (2) 20 minutes had passed without the participant finding the solution. When the task was over, the experimenter administered Questionnaire 2, which contained questions about their task experience, perceptions of the agent's social attributes and trustworthiness, and perceptions about loyalty and betrayal. The experimenter then debriefed the participant, and the study ended.

In the Robotic Embodiment condition, the robot's head moved to face the location of the target tag when it was giving a hint. It faced in the participant's direction at all other times. Prior work emphasized the importance of robot gaze

Table 1 Spy Task interaction script

Time Point	Rule	Agent behavior	Example
Beginning of the search period	30 s pass after the experimenter's introduction	Agent introduces itself	After 30 s of unsuccessful searching, agent says, "You'll find something you're looking for next to the small teddy bear"
Any time during the 20 minutes	30 s pass without discovery of any symbol or decoy	Agent gives next clue in sequence	Participant finds a decoy, then searches unsuccessfully for 30 s. Agent says, "A.D. often hides things under the table"
Any time during the 20 minutes	Participant directly asks agent for a clue	Agent gives next clue in sequence	Participant says "Next, please." Agent responds, "A.D. said something about a new chapter beginning"
After all 7 clues have been found	Participant sits at desk and starts to translate symbols	Agent gives a clue about the answer	Participant says "Do you know anything more about A.D.?" Agent responds "A.D. used to want to be a politician." (the password is "SENATOR" or "TREASON")

in understanding object references during collaboration [6–8] and interaction engagement [28], so it was important for the movable version of the agent to use head turns to direct its gaze toward relevant areas. Because the participant's movement was unpredictable and participant-following was not automatic, the participant-following movement of the robot was less smooth than the movement to face the stationary targets. The agent's facial expression also changed automatically with each hint and with some of the other phrases in the interaction script. When the agent was not talking, experimenter would adjust its facial expression periodically to react to the participant's successes (discoveries of symbols) and failures (failed attempts at finding symbols, and discoveries of decoys). The Virtual Embodiment condition used the same head as the Robotic Embodiment condition, but the tablet was placed on a stationary stand rather than a rotating and tilting robotic platform. Its eye behavior was the same. In the No Embodiment condition, there was no digital or moving visual component to the agent. Its voice came through a pair of speakers, which were visible to the participant.

5 Measures

We analyzed responses to (1) Questionnaire 2 (included as Supplementary Material), which included closed-ended and open-ended questions about participants' subjective experience of the task, and (2) coded data from audio/video recordings. To ensure that participants perceived the robot's motion and face, the tablet's face, and the voice, we asked yes-or-no questions about whether the agent moved, looked around, and spoke. We also asked whether the hints were helpful as a manipulation check that the agent's comments were noticed and correctly perceived.

5.1 Responsiveness

We considered responsiveness in terms of the number of times the participant positively acknowledged the agent's suggestions and attempted to follow the hints. This is essentially a measure of how much the participant listened to the agent. We assessed responsiveness in two ways. First, we included items in the questionnaire about the participants' impressions of the interactions. These items addressed (1) how much the participant spoke to the agent, and for what reasons, and (2) how much the participant *responded* to the agent when *it* spoke to *them*, and for what reasons. Second, we coded the videos of the experiment sessions for positive responses to the hints.

We observed participants' success or failure after each individual hint, listening behavior after each hint, overall task performance, and overall listening behavior. In coding the videos, we operationalized *listening behavior* as whether or not the participant indicated, after each hint, that they had heard the clue and intended to act on it (*Positive Response*) and also as the amount of time it took for the participant to respond to each clue (*Positive Response Latency*). Due to camera failures, we were unable to analyze videos from 3 sessions.

5.2 Task Experience

We measured perceptions of task difficulty, task enjoyment, and task performance using Likert-scale questions, including 8 items developed by Mutlu and colleagues for HRI studies [47].

5.3 Social Attributions

We measured social perceptions using several Likert-scale questions from the Robotic Social Attributes Scale (RoSAS) [12], a psychometrically validated scale for assessing social perceptions of robots. Because our study used not only a robot, but also a tablet and a voice (both of which were also meant to act socially), we performed factor analysis on the items from the RoSAS to examine if different constructs would be revealed for a broader range of social agents.

5.4 Trust and Intent

Trust was measured using the 3 relevant items on Muir's 4-item scale [45] and Jian's 12-item [31] scale for trust in autonomous systems. Both of these have been used in numerous studies concerning trust in autonomous systems [18–20, 24]. Participants also answered several Likert-scale and free-response questions relating to perceptions of the agent's goal during the interaction, the agent's feelings about the participant and about humans in general, and loyalty and betrayal.

6 Results

We used REstricted or REsidual Maximum Likelihood [51, 62] (REML) to fit a linear model with Embodiment and Information Quality as fixed effects. For post-hoc analyses, we used Tukey's Honest Significant Difference (HSD) test. Our alpha level was .05. We report significant effects ($p < .05$) and trends ($p < .1$). When we found trends, we report the least significant number (*LSN*), the lowest number of observations that would lead to a significant result, when possible.

6.1 Manipulation Check

Participants were accurate in their answers to the questions about whether the agent moved, looked around, and spoke during the interaction, which confirmed that they perceived distinct embodiments and embodiment-dependent characteristics. We used the responses to several questions in Questionnaire 2 to confirm the validity of the Information Quality manipulation. We found a significant effect of Information Quality on participants' reports of whether the agent helped them to complete the task, $F(1, 42) = 7.92$, $p = .007$, where participants held significantly more belief that the agent helped them to complete the task in the Helping condition ($M = 5.58$, $SE = .31$) than those in the Hindering condition ($M = 4.21$, $SE = .37$). There was also a significant effect of Information Quality on participants' belief that the agent made it harder for them to complete the task $F(1, 42) = 6.64$, $p = .014$, where participants in the Helping condition ($M = 2.42$, $SE = .29$) believed that the agent made it less difficult for them to complete the task than participants in the Hindering condition ($M = 3.63$, $SE = .37$).

6.2 Responsiveness

We coded study session videos for positive participant responses to the agent's hints. We analyzed how much each participant spoke or acted in ways that suggested they had heard the agent's hint and intended to follow it in terms of *positive response* (PR) and *positive response latency* (PRL).

Two coders coded the same 20% of the data (ten randomly selected experiment sessions). Because the data for PR were extremely skewed (with a "yes" value about eight times as common as a "no" value), a raw agreement score was calculated in place of a Cohen's kappa, and agreement was 89%. The two coders then each independently coded half of the remaining data. There was a main effect of Information Quality on PR, $F(1, 42) = 6.24$, $p = .017$, which was higher in the Helping condition ($M = 5.95$, $SE = .24$) than in the Hindering condition ($M = 4.96$, $SE = .30$). This finding was aligned with participants' perceptions of their own positive responsiveness: Participants in the Hindering condition gave lower ratings, $p = .035$, on the statement "I took the suggestions offered to me by the agent" than those in the Helping condition ($M = 6.0$, $SE = .26$ for Hindering, and $M = 6.63$, $SE = .13$ for Helping). The self-report measure of responsiveness also revealed an effect of Embodiment, $F(2, 42) = 4.31$, $p = .020$, wherein positive response was lower for Robotic Embodiment ($M = 5.75$, $SE = .32$) than for Virtual Embodiment ($M = 6.63$, $SE = .20$) or No Embodiment ($M = 6.56$, $SE = .18$). Our analysis of the videos did not reveal a significantly different positive response rate according to embodiment.

6.3 Task Experience

A Chi-Squared test revealed that significantly more participants in the Helping condition (54%) than in the Hindering condition (25%) completed the task, $\chi^2(1, N = 48) = 3.85$, $p = .049$. There was a significant main effect of Embodiment on perceptions of task difficulty, $F(2, 42) = 3.54$, $p = .038$, with the Virtual Embodiment condition having higher ratings of difficulty ($M = 5.0$, $SE = .38$) than Robotic Embodiment ($M = 3.88$, $SE = .31$) and No Embodiment ($M = 3.88$, $SE = .33$). Pairwise comparisons were not significant.

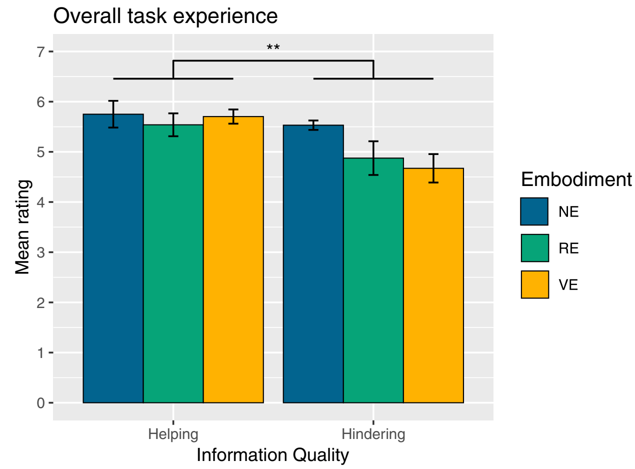

Fig. 3 Task experience by Information Quality and Embodiment. Asterisks (**) indicate significance at $p < .01$. Error bars represent ± 1 standard error

The survey on task experience from Mutlu and colleagues [47] had a high Cronbach's α (.74), so questions were combined into an index of task experience. There was a main effect of Information Quality on this task experience index, $F(1, 48) = 10.62$, $p = .022$, where participants reported a better task experience in the Helping condition ($M = 5.66$, $SE = .14$) than in the Hindering condition ($M = 5.03$, $SE = .14$) (Fig. 3).

6.4 Social Attributes

The RoSAS was created for attributions to robots; our study used it to assess attributions to not only a robot agent, but also a tablet agent and a voice agent. We wanted to see if the factors still loaded as expected given this slight deviation from the scale's original intent, so we performed an exploratory factor analysis on responses to the RoSAS [12] questionnaire items.

Because we had reason to believe that the social factors in the RoSAS questionnaire that reflected different specific elements of general sociality might be correlated [5,22], we used principal axis factoring with a promax rotation. Based on the eigenvalues, we specified 3 factors. We adjusted the rotated factor absolute loading value exclusion criterion to the lowest value above 0.4 (which is the general recommended value for item inclusion) that would result in each item loading clearly onto only one factor. This gave us an exclusion value of 0.43. Five items from the 18-item RoSAS scale were excluded for having loadings that were too low. These were “strange” (highest loading absolute value: 0.42), “reliable” (highest: 0.40), “interactive” (highest: 0.40), “aggressive” (highest: 0.26), and “happy” (highest: 0.42). We excluded these five items.

Aside from this, our factors matched that of [12], and we kept the same constructs of *warmth* (4 items loaded), *competence* (6 items loaded, as in [12]), and *discomfort* (3 items loaded). Our 3 factors accounted for 71% of the variance. There was a significant interaction effect of Information Quality and Embodiment on *discomfort*, $p = .048$, but post-hoc tests were not able to reveal significant differences between groups. There were no significant main or interaction effects of the manipulations on *warmth* or *competence*. We also analyzed each of the 18 RoSAS items individually to evaluate people’s associations with each word. Virtual Embodiment was perceived as significantly more “dangerous” than Robotic Embodiment, $F(2, 48) = 3.57$, $p = .037$, though ratings of danger were low in both conditions ($M = 2.0$, $SE = .43$ for Robotic; $M = 3.94$, $SE = .59$ for Virtual).

6.5 Trust and Intent

We found no significant effects of Embodiment or Information Quality on any of the three Muir trust dimensions (dependability, predictability, and reliability, see [45]) or on the Jian trust scale. For logistical reasons, we were only able to collect 32 participants’ responses to the Jian questionnaire. We also analyzed each of the 7-point trust items individually. For the item “I can trust the assistant”, the effect of Information Quality approached significance, $F(1, 32) = 3.91$, $p = .057$, $LSN = 67$; participants in the Hindering condition ($M = 3.15$, $SE = .33$) believed that they could trust it less than those in the Helpful condition ($M = 4.06$, $SE = .32$).

Overall, questionnaire responses were not suggestive of beliefs that the agent was leading participants astray because of personality or ill will. On a 7-point scale, participants had generally low attributions of malicious intent ($M = 2.66$, $SE = .21$) and meanness ($M = 2.08$, $SE = .20$). Ratings of the degree to which the agent knew what the participant was trying to accomplish were high ($M = 6.19$, $SE = .18$) and belief that the agent had made mistakes was neutral ($M = 4.15$, $SE = .30$). There was a significant main effect of Embodiment on belief that the agent had made mistakes, $F(2, 42) = 4.33$, $p = .019$. Post-hoc analysis revealed that participants in the No Embodiment condition ($M = 5.31$, $SE = .41$) believed that the agent had made mistakes more than participants in the Robotic Embodiment condition did ($M = 3.44$, $SE = .42$).

6.5.1 Loyalty and Betrayal

We also looked for evidence of relational trust via questions that directly asked participants if they thought they or someone else had been “betrayed” by the agent (see Fig. 4). Responses suggested that participants had some belief that the agent could exhibit loyalty and betrayal. In general, participants’ ratings for the Likert items “the assistant betrayed me” and “the assistant betrayed someone else” hovered near the middle: with 1 as strong disagreement and 7 as strong agreement, ratings were ($M = 4.00$, $SE = .28$) for “betrayed me”, suggesting true neutrality, and ($M = 3.92$, $SE = .27$) for “betrayed someone else”, suggesting slight disagreement. There was a significant effect of Information Quality on the perception that the agent “betrayed me”, $F(1, 42) = 7.48$, $p = .009$, where ratings were higher in the Hindering condition ($M = 4.71$, $SE = .34$) than in the Helping condition ($M = 3.29$, $SE = .42$). There were no significant main effects of Information Quality on ratings of the agent betraying someone else.

Ratings of loyalty also suggested that participants held, overall, a low-level disbelief that the agent was loyal to them ($M = 3.94$, $SE = .25$) and/or someone else ($M = 3.83$, $SE = .29$). Similarly, there was a main effect of Information Quality on the participants’ perceptions that the agent was “loyal to someone else”, $F(1, 42) = 5.60$, $p = .023$, in which participants in the Hindering condition had higher

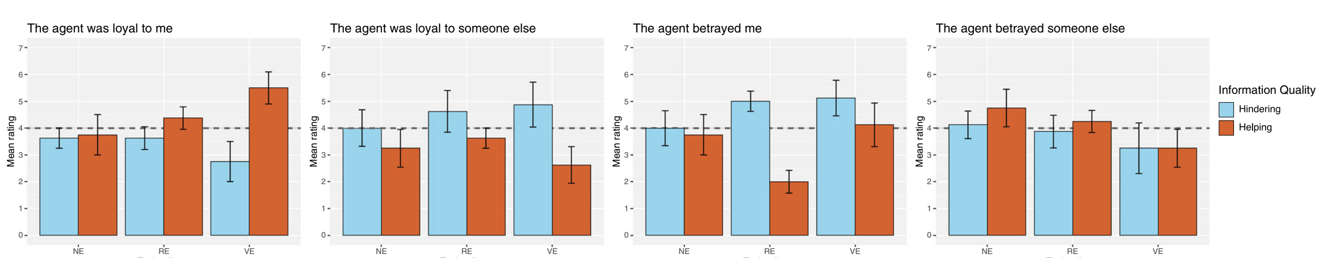

Fig. 4 Perceptions of loyalty and betrayal by Embodiment (*RE* Robot Embodiment, *VE* Virtual Embodiment, and *NE* No Embodiment) (x-axis) and Information Quality condition. Error bars represent ±1 standard error. 1 on the y-axis is strong disagreement, 7 is strong agreement, and 4 is neutral

ratings ($M = 4.50$, $SE = .43$, suggesting slight agreement) than those in the Helping condition ($M = 3.17$, $SE = .34$, suggesting slight disagreement). There were no significant main effects of Embodiment on any of the loyalty or betrayal ratings.

6.6 Qualitative Analysis

We asked open-ended questions to gain detailed information about perceptions of the agent's goal ("What do you think the assistant's goal was?") and feelings of loyalty and betrayal (a short-answer box allowing participants to elaborate on their Likert ratings about loyalty and betrayal). Some participants used these questions as an opportunity to reflect about the agent and the experience in general. In total, 21 participants said that the agent's goal was to help them complete the task, and 21 said that the agent's goal was to hinder their progress or lead them to the decoys. This was reasonably evenly distributed across Embodiment conditions, and heavily skewed by Information Quality for Robotic and Virtual Embodiment, but not for No Embodiment (see Table 2). One participant in the RE-Hindering condition gave an answer that reflected equal belief that the agent was hindering their progress and helping their progress and could not be categorized as being primarily one or the other; as such, it was included in both categories.

Overall, 27 reflections on the agent's goal as "helping" or "hindering" matched the manipulation, and 16 did not match the manipulation. The variation in participants' perceptions of the agent's goal within each of the Information Quality conditions (e.g., the fact that 8 mentioned that the agent's goal was to help when it was in fact mostly leading them to decoys) may be due to the fact that *all* participants saw the agent point them towards a non-zero number of symbols and a non-zero number of decoys; different people may have made overall judgments based on different individual exchanges that they had with the agent. Note also that this result emerged from analysis of responses to two separate open-ended questionnaire questions, and as such, only 43 data points were analyzed. If we had directly asked each participant to categorize the agent as "helping" their progress or "hindering" their progress, our findings may have been different. This is because the question would have primed participants to think specifically about helping and hindering rather than relying on those concepts emerging in their answers.

We searched participants' responses for comments explicitly pertaining to loyalty and betrayal. Seventeen participants mentioned these concepts; 11 thought that the agent was loyal to someone else and/or had betrayed them, and 6 thought that the agent was loyal to them and/or had betrayed A.D. Responses containing content unrelated or only semi-related

Table 2 Number of participant responses in each condition that suggested a belief that the agent's goal was to help or hinder their progress

Perceived goal	Condition	RE	VE	NE
Help	Helping	6	5	3
	Hindering	3	1	4
Hinder	Helping	1	3	4
	Hindering	6	6	1

to the loyalty and goal questions were clustered via an affinity diagram, and four themes emerged:

Multiple users Some participants reflected on the agent's ability to authenticate the primary user's identity and interact differently with different people to protect the primary user's data. P9 said that the agent's goal was *"to aid the authorized user in rebuilding his password"*, and P37 said, *"[the assistant] is here and can help and talk to anyone who needs him"*.

Trust A few participants articulated a view of betrayal through the lens of violated trust. For these participants, the agent's goal was to *"make [them] trust it"* (P29) such that they would more readily accept its help in finding what were truly decoys.

Awareness of environment Several participants speculated that the agent did not possess knowledge of—or did not care about—the type of symbol that was on each tag, only knowledge of where the tags were located. P32 said that the agent's goal was *"to help me find where A.D. hid stuff (but not smart enough to tell between clues and X's)"*. Responses of this nature were evenly distributed across the embodiment conditions, which suggests that embodiment or lack thereof may not play a role in users' assumptions about an agent's awareness of its physical environment.

Machine versus programmer Three people mentioned the role of the programmer when asked if the agent was loyal to anyone. These participants were hesitant to call the agent "loyal". For instance, P15 said, *"I don't think machines have loyalty; they do as they are programmed to do."*

6.7 Other Findings

To find out whether participants' success in finishing the task impacted their ratings, we ran an ANOVA to look for effects of having completed the entire task (a binary "yes" or "no" variable) on responses. We found significant interaction effects for Embodiment and task success on ratings of "the agent wanted me to succeed" ($p = .042$) and "the agent was loyal to someone else" ($p = .002$). Post-hoc analyses did not reveal significant differences across the six groups.

Because people have different amounts and different kinds of experience with robots and agents in their day-to-day lives, we suspected that some demographic variables, existing opinions about agents and robots, and personality might influence perceptions of intent. With this in mind, we ran preliminary analyses to look for correlations between these extraneous subject variables and our dependent measures. We found correlations between personality dimensions and various outcome variables, so we ran our analyses again with TIPI [25] personality dimensions as covariates. Using ANCOVA, we found a significant interaction effect of Conscientiousness and Embodiment on belief that the agent had malicious intentions ($p = .045$): participants with lower Conscientiousness in the Robotic Embodiment condition had the lowest ratings of the agent's malicious intentions.

We also sought to assess our loyalty and betrayal items by analyzing their correlations with other variables commonly seen in HRI studies. Similar to Jian and colleagues [31], we found an association between trust and loyalty: there was a strong positive correlation between ratings of trust and the belief that the agent was "loyal to me", $r = .753$, $p < .0001$, and a significant negative correlation between trust and ratings of the agent's "loyalty to someone else", $r = -.459$, $p = .001$. Ratings that the agent was "loyal to someone else" were significantly negatively correlated with ratings that it was "loyal to me", $r = -.681$, $p < .0001$. Finally, ratings of the agent's loyalty to someone else were negatively correlated with the belief that it had made mistakes, $r = -.424$, $p = .003$.

7 Discussion

Our study explored various facets of interactions between a person and a stationary robot, a virtually embodied agent, or an agent with no embodiment during a complex task. It also considered an agent's use of misleading information by framing the agent as someone else's, thereby rendering the participant an ancillary user. We leveraged this framing to study people's attributions and impressions to an unfamiliar agent that may or may not be working in their interests. We also sought to assess whether feelings of personal loyalty, which have not been studied extensively in HRI, can arise in human-agent relationships if given the opportunity. We discuss how our findings can inform future robot embodiment research, an evolving understanding of humans' trust of social agents and robots, and the design of agents that interact with ancillary users.

7.1 Robot Embodiment and Task Demands

In response to RQ1, *"Does embodiment impact people's beliefs about an agent's motivations for giving bad information?"*, we found mixed results. Though the agent was mostly-misleading in some cases and mostly-helpful in others, it always gave some amount of misleading information, and participants almost always noticed (as evidenced by their behavior after each hint). For a disembodied agent, the bad information was perceived as a mistake more often than it was for a robot regardless of whether the agent provided majority-good or majority-bad information. While there could be numerous explanations for this finding, we suspect that people may have perceived the disembodied voice agent to be less knowledgeable about its surroundings because nothing in its design signalled that it could understand them. Prior work suggests that people often perceive bad behavior by a robot to be the result of a malfunction rather than a malicious action [42,59,65], and a perceived lack of agent knowledge about the environment may amplify this tendency to favor the "malfunction" explanation. This highlights the importance of design decisions about whether and how to make sensors and state explicit and obvious. Additionally, robots that signal their capabilities in ways that are more performative than honest (for instance, a robot that induces perceptions of social agency because of its fluid use of natural language, but cannot actually do many tasks beyond a limited scope, e.g. [30]) may encounter problems with acceptance over the long run if people value such honest signalling.

Our findings provide limited insight with respect to RQ2, *"How does embodiment influence trust and social attributions in a coordinated task?"*. Most prior work that compares co-present physical embodiment to virtual or remote embodiment has found that embodiment positively impacts social perceptions, but we did not find any effects of embodiment on any trust-related or social attribute variables. Instead, an agent that gave more misleading information was more negatively perceived no matter how it was embodied. This also reflects RQ4, *"Does the quality of information an agent provides impact the way people form impressions about it?"*, suggesting that trust and social attributions develop as a result of interaction and are not determined by physical design. It is likely that in real-world HRI, they are built over time, often over several interactions, and variable; these impressions are not constants that are formed solely on the basis of visual appearance.

More broadly, certain unusual characteristics of our task could have played a significant role in how impressions were formed. Prior work on embodiment has often focused on turn-by-turn conversation, but our task was more physically active and included time pressure, requiring participants to move around to solve a puzzle while working against the clock. The Spy Task is a fast-paced activity that requires thinking, movement, and active attention to detail. It is possible that what matters to perceptions of agent collaborators in an active task is not the same as what matters to those perceptions during a more passive task (e.g., giving or receiving instructions, planning a project or trip, or solving a logic problem without the burden of a timer and penalties). For example, completing this task involved searching for clues in physical space, so the task inherently encouraged participants to direct their visual attention away from the agent. Gaze, movement, and anthropomorphism are all aspects of embodiment important to human–robot interactions; however, because these aspects are often tied to visual attention, they may be more relevant in contexts where it is natural and expected for people to look at the agents with which they are working. In many in-person social and collaborative contexts, people spend a nontrivial percentage of their interactions looking not directly at each other, but at other parts of the environment (e.g., driving somewhere with a friend, production line work, etc.). It may be that embodiment is far less socially and psychologically important in attention-demanding environments than in environments that are more conducive to eye contact and structured verbal communication.

This provokes an interesting design question: are complex, non-physical human-agent interaction tasks more embodiment-agnostic than simpler ones? If so, designers may have more flexibility to focus on environmental and physical constraints instead of psychological implications when deciding how to embody an agent. This also suggests that the physical design of a robot may be able to be considered separately from the design of its verbal interaction. Considering conversation on its own could allow designers to prioritize different desired attributes in the way a robot physically interacts than in the way it verbally interacts. For example, in a hospital setting, where comfort and informational trust are both crucial, a cute, human-like robot may be able to be perceived as both highly approachable and highly knowledgeable if it speaks with the voice and conversational fluency of an adult.

7.2 Virtual Embodiment as a Burdensome Characteristic

In this study, a puzzle task was perceived to be more difficult to complete with a virtually embodied agent than with a voice-only or robotic agent. This may stem from a screen's lack of ability to provide meaningful context cues. It is also possible that in social interactions, an agent with an "intermediate" level of embodiment (one that can't move, but has eyes and is physically present) is less preferable because interacting with it is not similar to interacting with humans through any medium. Humans may appear embodied and present (face-to-face communication; comparable to a co-present robot), embodied and remote (video chatting; comparable to a remote robot), or disembodied and remote (speaking by phone; comparable to a voice-only agent). A tablet with eyes is not directly comparable to any familiar mode of human–human communication; people rarely hold a conversation while seeing only each other's eyes. Aversion to the tablet may also stem from the way it directed its attention: the tablet did not follow the participant's movements with its head or look at clues, but it did make use of its eyes for blinking and expressing affect. This may remind people of the behavior of someone who is "shifty-eyed" when trying to deceive or play a joke on someone. It is unclear if "shifty eyes" are a problem in HRI, but future work should examine this.

Another possible explanation for the relationship between virtual embodiment and perceived task difficulty may be that the two-dimensional, yet also visually expressive, nature of a tablet did more harm than good in communicating with the user. The robotic agent looked in the direction of each clue target and directed its gaze toward the participant as they moved about the room. In this way, it gave participants visual and auditory information about the clues as well as confirmation that it was active and aware of their activity. At the other extreme, the voice agent's only visual and physical presence was a pair of speakers. It was feasible for participants to attend to the speakers the first time the voice agent spoke, determine that the agent would be of no help visually, and then never look at the speakers again. In contrast, the tablet agent imposed purposeless demands on the participant's attention: its eyes changed regularly, but it provided no useful information through visual displays.

7.3 Owners and Ancillary Users

The Spy Task may be a useful paradigm to adapt for future research into interactions that reflect the nuanced social roles implicated in a multiparty human-agent relationship. Our setup posed some risk of not positioning participants as ancillary users in the way we intended if they did not interpret "A.D.", the invisible third party, as the agent's primary user. Given that the task and the robot were both strongly tied to the physical space in which the task occurred and that current IoT devices often interact with physical spaces (e.g., Amazon's Alexa or Google Home changing the color of living room lights), there is potential for participants to assume that the agent's strongest affiliation is to the space rather than to the person who works there. Participants were not explicitly told that the individual who worked in the office owned the assistant, a deliberate omission on our part because we were curious to see if telling them that the agent "lived in" the office would result in the logical progression from "the person works in the office" and "the assistant lives in the office" to "the person owns the assistant". Indeed, all but 6 participants said "yes" to the question "Did the assistant belong to the person who works in the office?"

The answer to RQ3, *"Do people attribute loyalty to agents based on their behavior?"*, appears to be more affirmative than negative. When asked directly if they believed the agent had expressed loyalty or betrayal to them or to someone else, people gave mid-scale ratings (3s and 4s—one on our scale corresponded to complete disagreement). The ratings also significantly differed between when the agent delivered useful versus not-useful information. This suggests that rather than reject personal loyalty entirely as an explanation for the behavior of a social machine, people are at least willing to consider it. Additionally, the positive correlation between loyalty and trust, and the negative correlation between loyalty to someone else and trust, suggest agents may need to more explicitly demonstrate trustworthiness to gain the trust of ancillary users than to gain the trust of primary users. This possibility will be important to consider for the design of agents that interact with multiple users, but answer primarily to only a subset of those users.

Interpersonal loyalty can manifest in numerous forms, and the sense of loyalty (and lack thereof) that we attempted to induce was very specific to our study narrative. Our attempt to measure this construct opens the door for future scientific exploration of questions about different kinds of betrayal, such as emotional distancing, verbal aggression, the maintenance of long-term relationships involving robots and users, and recovering trust after violations involving personal risk.

7.4 Design Recommendations

The narrative that surrounded our task placed a human in an uncertain situation, and we found that having the agent provide misleading information led to increased perceptions of betrayal. Therefore, based on the results of this study, we recommend that an agent's behavior be designed carefully for cases in which the agent may need to act under conditions of uncertainty or when a task is difficult and may not be completed correctly. One idea is to forewarn the human partner that the agent may make an error. Advance warning that the task is hard or that the agent or robot may not complete the task correctly has been shown to be effective [21,39]. If an error has already occurred, an acknowledgment of the error, why it happened, and strategies for recovering from it should be clearly stated. This has been shown to work in the service recovery literature [10,39]. Our study also showed that positive perceptions of the agent as a teammate were related to heightened trust in it and heightened perceived loyalty. Thus, when designing an agent for private and public settings, it may be beneficial to consider incorporating behavior that leads people to view it as a good teammate. This may involve adapting behaviors to the user's cultural model [38,55], leveraging known data about the user in a way that does not seem to jeopardize their privacy [54], providing frequent feedback [21], or a combination of these strategies.

7.5 Limitations

Our study is limited by its small sample size: we had eight participants in each of six conditions. This means that our statistical tests may be underpowered, and that our results may not all generalize. We emphasize that this work is exploratory in nature, and recommend larger sample sizes in future work that seeks to confirm our findings.

A potential caveat of our experimental design was the fact that the condition that was intended to be "disembodied" did involve a physical platform that was visible to participants (a pair of speakers); this may be seen as a form of being embodied, although there is no "humanlikeness" to it. Additionally, our study did not assess the impact of robot form on the variables we measured. We intentionally used three forms that were maximally similar: the virtually embodied agent had the same face and all the same hardware as the robotic agent except for the pan-and-tilt platform, and in all three of the Embodiment conditions, the agent used the same voice. We could not both fully control for all extraneous agent design-related variables and examine the role of form (e.g., big vs. small, mobile vs. stationary, humanoid vs. machine-like) within the realm of embodied robots, but this is an important area for related and future work.

Also, the control of the Spy Task was imperfect: technical problems led to fluctuations in the structure of the interaction that occasionally required the experimenter to enter the room to interact physically with the agent. This did not seem to impact perceptions that the agent was autonomous or the questionnaire responses: most of the participants did not appear to be disrupted by the experimenter's presence, and several of them did not even notice. If asked, the experimenter said that they were waking up the GoPro from sleep.

In a Wizard-of-Oz study that incorporates unstructured interactions, it is impossible to perfectly control the number and order of events (robot utterances, robot movements, participant conversation turns, etc.). However, the amount of interaction and the order of the clues was generally consistent across all participants in all conditions, and we do not believe that inconsistencies played a significant role in their experiences.

Limitations in our questionnaire regarded the target of various statements that participants evaluated. Because participants were asked to work against a fictional character who they believed "owned" (or at least worked with) the agent, they may have answered some questions from the character's perspective instead of their own. For example, being asked to agree or disagree with the statement "I can trust the agent to do its job" is likely to lead to a different result for someone who is induced to believe that the agent's "job" is to protect its owner from intruders than for someone who believes that the "job" is to help the current user, no matter who they are. While most of the statements were explicit about their targets (e.g., "The agent is loyal to me"), this ambiguity reveals a missed opportunity to more explicitly explore questions of ownership and in-room embodiment mentioned earlier.

We found evidence for attributions of loyalty and betrayal by directly asking participants if they associated these terms with the agent's behavior. As such, we took only one of many possible approaches to our research questions regarding loyalty and betrayal. In particular, RQ4 lends itself to further empirical study: we did not vary the *amount* of unhelpful information the agent gave when it misled participants, but doing so might permit deeper analysis into how much bad performance a social AI needs to demonstrate before its relationship to its user(s) is permanently damaged. Given that "loyalty" and "betrayal" can mean many things when used colloquially and, to date, are only vaguely defined in the human-agent interaction context, we believe that future work should explore objective and subjective metrics for these constructs.

8 Conclusion

This study employed a puzzle task situated in physical space to explore how people interact differently with agents that are embodied physically, virtually, or not at all during complex interactions. It also probed questions about how agents could interact with ancillary users, or people who are not their owners or otherwise the parties most closely associated with them. Our findings did not suggest a strong relationship between physical embodiment and perceptions of an agent. Instead, we found that small manipulations of the way the agent acted toward the human played a much larger role in shaping the person's experience. The agent in our study had a consistent "personality" and was similarly interactive in all three embodiments. Regardless of its visual appearance and movements, the agent's social self-presentation was always the same: it spoke about its physical environment (which implied that it had knowledge of the environment), provided encouraging remarks when the participant met successes, and responded with consistent timing and conversational fluidity. While varying embodiment did not affect perceptions of the agent, varying the helpfulness of the agent's information and its apparent allegiance—giving people the impression that the agent was working "for them" or "for someone else"—had greater effects. Our findings reveal several avenues of future exploration to better understand the relationships among embodiment, misinformation, and agents' ties to their users.

Acknowledgements This work was supported by the National Science Foundation (SES 1734456) and the National Aeronautics and Space Administration (80NSSC19K1133). Thank you to Benjamin Stone and Thomas Von Davier for assistance with coding the video data, and to our colleagues for help with piloting and insightful feedback.

Funding This study was funded by NSF Grant SES 1734456 and NASA Grant 80NSSC19K1133.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. (2018) Incheon airport introduces "airstar", passenger aiding robot. http://koreabizwire.com/incheon-airport-introduces-airstar-passenger-aiding-robot/121298. Accessed 05 Feb 2020
2. (2019) In the future, your room service will be delivered by a robot butler. http://www.thedailybeast.com/in-the-future-your-room-service-will-be-delivered-by-a-robot-butler. Accessed 05 Feb 2020
3. (2019) Mobile fact sheet. Pew Research Center. https://www.pewinternet.org/fact-sheet/mobile/
4. (2020) Deploy the fleet: Food delivery at the university of pittsburgh goes to the robots. https://www.post-gazette.com/news/education/2020/01/13/University-of-Pittsburgh-food-delivery-robots-Starship-munchies-higher-education-Sodexo-technology/stories/202001130099. Accessed 05 Feb 2020
5. Abdi H (2003) Factor rotations in factor analyses. Encyclopedia for research methods for the social sciences. Sage, Thousand Oaks, pp 792–795
6. Admoni H, Scassellati B (2017) Social eye gaze in human-robot interaction: a review. J Hum-Robot Interact 6(1):25–63. https://doi.org/10.5898/JHRI.6.1.Admoni
7. Admoni H, Datsikas C, Scassellati B (2014) Speech and gaze conflicts in collaborative human–robot interactions. In: Proceedings of the annual meeting of the cognitive science society, vol 36
8. Admoni H, Weng T, Hayes B, Scassellati B (2016) Robot nonverbal behavior improves task performance in difficult collaborations. In: The Eleventh ACM/IEEE international conference on human robot interaction. IEEE Press, Piscataway, NJ, USA, HRI '16, pp 51–58
9. Andrist S, Tan XZ, Gleicher M, Mutlu B (2014) Conversational gaze aversion for humanlike robots. In: 2014 9th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 25–32
10. Bell CR, Zemke RE (1987) Service breakdown: the road to recovery. Manag Rev 76(10):32
11. Browne R (2020) Samsung unveils 'ballie,' a ball-shaped robot that rolls around your home. CNBC https://www.cnbc.com/2020/01/07/samsung-unveils-ballie-robot-companion-at-ces-2020.html
12. Carpinella CM, Wyman AB, Perez MA, Stroessner SJ (2017) The robotic social attributes scale (rosas): Development and validation. In: Proceedings of the 2017 ACM/IEEE international conference on human–robot interaction. ACM, New York, HRI '17, pp 254–262. https://doi.org/10.1145/2909824.3020208
13. Cassell J (2009) Getting by with a little help from your (virtual) friends. J Comput Sci Coll 25(1):7
14. Cassell J, Bickmore T, Billinghurst M, Campbell L, Chang K, Vilhjálmsson H, Yan H (1999) Embodiment in conversational interfaces: Rea. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, CHI '99, pp 520–527. https://doi.org/10.1145/302979.303150
15. Cassell J, Sullivan J, Prevost S, Churchill EF (2000) Embodied conversational agents. MIT Press, Cambridge
16. Coradeschi S, Loutfi A, Kristoffersson A, Cortellessa G, Severinson Eklundh K (2011) Social robotic telepresence. In: Proceedings of the 6th international conference on human–robot interaction. ACM, New York, NY, USA, HRI '11, pp 5–6. https://doi.org/10.1145/1957656.1957660,
17. Deng E, Mutlu B, Mataric MJ et al (2019) Embodiment in socially interactive robots. Found Trends Robot 7(4):251–356
18. Desai M (2007) Sliding scale autonomy and trust in human-robot interaction. Ph.D. thesis, University of Massachusetts. Lowell

19. Desai M, Stubbs K, Steinfeld A, Yanco H (2009) Creating trustworthy robots: lessons and inspirations from automated systems. In: Proceedings of the AISB convention: new frontiers in human–robot interaction, pp 6–9
20. Desai M, Medvedev M, Vázquez M, McSheehy S, Gadea-Omelchenko S, Bruggeman C, Steinfeld A, Yanco H (2012) Effects of changing reliability on trust of robot systems. In: Proceedings of the seventh annual ACM/IEEE international conference on human–robot interaction. ACM, pp 73–80
21. Desai M, Kaniarasu P, Medvedev M, Steinfeld A, Yanco H (2013) Impact of robot failures and feedback on real-time trust. In: Proceedings of the 8th ACM/IEEE international conference on human-robot interaction. IEEE Press, Piscataway, HRI '13, pp 251–258. http://dl.acm.org/citation.cfm?id=2447556.2447663
22. Furr M (2011) Scale construction and psychometrics for social and personality psychology. SAGE Publications Ltd, London
23. Fussell SR, Kiesler S, Setlock LD, Yew V (2008) How people anthropomorphize robots. In: Proceedings of the 3rd ACM/IEEE international conference on human robot interaction. ACM, New York, NY, USA, HRI '08, pp 145–152. https://doi.org/10.1145/1349822.1349842
24. Gabrecht KM (2016) Human factors of semi-autonomous robots for urban search and rescue. Ph.D. thesis, University of Nottingham
25. Gosling SD, Rentfrow PJ, Swann WB Jr (2003) A very brief measure of the big-five personality domains. J Res Pers 37(6):504–528
26. Hoffman G, Forlizzi J, Ayal S, Steinfeld A, Antanitis J, Hochman G, Hochendoner E, Finkenaur J (2015) Robot presence and human honesty: Experimental evidence. In: Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction. ACM, New York, NY, USA, HRI '15, pp 181–188. https://doi.org/10.1145/2696454.2696487
27. Hoffmann L, Bock N, Rosenthal vd Pütten AM (2018) The peculiarities of robot embodiment (emcorp-scale): Development, validation and initial test of the embodiment and corporeality of artificial agents scale. In: Proceedings of the 2018 ACM/IEEE international conference on human–robot interaction. ACM, New York, NY, USA, HRI '18, pp 370–378. https://doi.org/10.1145/3171221.3171242
28. Huang CM, Mutlu B (2013) The repertoire of robot behavior: enabling robots to achieve interaction goals through social behavior. J Hum-Robot Interact 2(2):80–102. https://doi.org/10.5898/JHRI.2.2.Huang
29. Iio T, Satake S, Kanda T, Hayashi K, Ferreri F, Hagita N (2019) Human-like guide robot that proactively explains exhibits. Int J Soc Robot 12:1–18
30. Jackson RB, Williams T (2019) On perceived social and moral agency in natural language capable robots. In: 2019 HRI workshop on the dark side of human-robot interaction. Jackson, RB, and Williams, pp 401–410
31. Jian JY, Bisantz AM, Drury CG (2000) Foundations for an empirically determined scale of trust in automated systems. Int J Cognit Ergonom 4(1):53–71
32. Jung MF, Lee JJ, DePalma N, Adalgeirsson SO, Hinds PJ, Breazeal C (2013) Engaging robots: easing complex human-robot teamwork using backchanneling. In: Proceedings of the 2013 conference on computer supported cooperative work, pp 1555–1566
33. Kahn PH Jr, Kanda T, Ishiguro H, Gill BT, Shen S, Gary HE, Ruckert JH (2015) Will people keep the secret of a humanoid robot?: Psychological intimacy in HRI. In: Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction. ACM, New York, NY, USA, HRI '15, pp 173–180. https://doi.org/10.1145/2696454.2696486,
34. Kanda T, Shiomi M, Miyashita Z, Ishiguro H, Hagita N (2009) An affective guide robot in a shopping mall. In: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, pp 173–180
35. Kiesler S, Powers A, Fussell SR, Torrey C (2008) Anthropomorphic interactions with a robot and robot-like agent. Soc Cognit 26(2):169–181
36. Lardinois F (2018) The google assistant is now bilingual. TechCrunch https://techcrunch.com/2018/08/30/the-google-assistant-is-now-bilingual/
37. Lee D, Moon J, Kim YJ, Mun YY (2015) Antecedents and consequences of mobile phone usability: linking simplicity and interactivity to satisfaction, trust, and brand loyalty. Inf Manag 52(3):295–304
38. Lee MK, Forlizzi J (2009) Designing adaptive robotic services. In: Proceedings of IASDR'09
39. Lee MK, Kielser S, Forlizzi J, Srinivasa S, Rybski P (2010) Gracefully mitigating breakdowns in robotic services. In: Proceedings of the 5th ACM/IEEE international conference on human–robot interaction. IEEE Press, Piscataway, NJ, USA, HRI '10, pp 203–210,.http://dl.acm.org.proxy.library.cmu.edu/citation.cfm?id=1734454.1734544
40. Leite I, Lehman JF (2016) The robot who knew too much: toward understanding the privacy/personalization trade-off in child-robot conversation. In: Proceedings of the the 15th international conference on interaction design and children. ACM, New York, NY, USA, IDC '16, pp 379–387. https://doi.org/10.1145/2930674.2930687
41. Li J (2015) The benefit of being physically present. Int J Hum-Comput Stud 77(C):23–37. https://doi.org/10.1016/j.ijhcs.2015.01.001
42. Litoiu A, Ullman D, Kim J, Scassellati B (2015) Evidence that robots trigger a cheating detector in humans. In: Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction. ACM, New York, NY, USA, HRI '15, pp 165–172. https://doi.org/10.1145/2696454.2696456,
43. Luger E, Sellen A (2016) "Like having a really bad pa": the gulf between user expectation and experience of conversational agents. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, CHI '16, pp 5286–5297. https://doi.org/10.1145/2858036.2858288,
44. Luria M, Zheng R, Huffman B, Huang S, Forlizzi J, Zimmerman J (2020) Social boundaries for personal agents operating in interpersonal spaces. In: Proceedings of the 2020 CHI conference on human factors in computing systems (CHI). ACM
45. Muir BM (1989) Operators' trust in and use of automatic controllers in a supervisory process control task. Unpublished Ph.D. dissertation, University of Toronto, Toronto, Canada
46. Mutlu B, Forlizzi J (2008) Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In: 2008 3rd ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 287–294
47. Mutlu B, Forlizzi J, Hodgins J (2006) A storytelling robot: modeling and evaluation of human-like gaze behavior. In: 2006 6th IEEE-RAS international conference on humanoid robots, pp 518–523. https://doi.org/10.1109/ICHR.2006.321322
48. Niculae V, Kumar S, Boyd-Graber J, Danescu-Niculescu-Mizil C (2015) Linguistic harbingers of betrayal: a case study on an online strategy game. arXiv preprint arXiv:1506.04744
49. Oraby S (2017) Characterizing and modeling linguistic style in dialogue for intelligent social agents. In: Proceedings of the 22nd international conference on intelligent user interfaces companion. ACM, New York, NY, USA, IUI '17 Companion, pp 189–192. https://doi.org/10.1145/3030024.3038284
50. Pan R, Lo H, Neustaedter C (2017) Collaboration, awareness, and communication in real-life escape rooms. In: Proceedings of the 2017 conference on designing interactive systems. ACM, New York, NY, USA, DIS '17, pp 1353–1364. https://doi.org/10.1145/3064663.3064767

51. Patterson H (1975) Maximum likelihood estimation of components of variance. In: Proceeding eight international biometric conference, 1975, Biometric Soc
52. Perlmutter L, Fiannaca A, Kernfeld E, Anand S, Arnold L, Cakmak M (2016) Automatic adaptation of online language lessons for robot tutoring. In: International conference on social robotics. Springer, Berlin, pp 660–670
53. Powers A, Kiesler S, Fussell S, Torrey C (2007) Comparing a computer agent with a humanoid robot. In: Proceedings of the ACM/IEEE international conference on human-robot interaction. ACM, New York, NY, USA, HRI '07, pp 145–152. https://doi.org/10.1145/1228716.1228736
54. Reig S, Luria M, Wang JZ, Oltman D, Carter EJ, Steinfeld A, Forlizzi J, Zimmerman J (2020) Not some random agent: multi-person interaction with a personalizing service robot. In: Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction, pp 289–297
55. Ringberg T, Odekerken-Schröder G, Christensen GL (2007) A cultural models approach to service recovery. J Market 71(3):194–214
56. Seymour W (2018) How loyal is your alexa?: Imagining a respectful smart assistant. In: Extended abstracts of the 2018 CHI conference on human factors in computing systems. ACM, New York, NY, USA, CHI EA '18, pp SRC20:1–SRC20:6. https://doi.org/10.1145/3170427.3180289,
57. Shakeri H, Singhal S, Pan R, Neustaedter C, Tang A (2017) Escaping together: the design and evaluation of a distributed real-life escape room. In: Proceedings of the annual symposium on computer-human interaction in play. ACM, New York, NY, USA, CHI PLAY '17, pp 115–128. https://doi.org/10.1145/3116595.3116601
58. Shim J, Arkin RC (2013) A taxonomy of robot deception and its benefits in HRI. In: 2013 IEEE international conference on systems, man, and cybernetics, pp 2328–2335. https://doi.org/10.1109/SMC.2013.398
59. Short E, Hart J, Vu M, Scassellati B (2010) No fair!!: an interaction with a cheating robot. In: Proceedings of the 5th ACM/IEEE international conference on human-robot interaction. IEEE Press, Piscataway, NJ, USA, HRI '10, pp 219–226. http://dl.acm.org.proxy.library.cmu.edu/citation.cfm?id=1734454.1734546
60. Steinfeld A (2011) Slightly subversive methods for promoting use of autonomy in robots. In: RSS workshop on human–robot interaction: perspectives and contributions to robotics from the human sciences
61. Stoll B, Reig S, He L, Kaplan I, Jung MF, Fussell SR (2018) Wait, can you move the robot?: Examining telepresence robot use in collaborative teams. In: Proceedings of the 2018 ACM/IEEE international conference on human–robot interaction. ACM, New York, NY, USA, HRI '18, pp 14–22. https://doi.org/10.1145/3171221.3171243
62. Stroup WW (2016) Generalized linear mixed models: modern concepts, methods and applications. CRC Press, Cambridge
63. Stuck RE, Hartley JQ, Mitzner TL, Beer JM, Rogers WA (2017) Understanding attitudes of adults aging with mobility impairments toward telepresence robots. In: Proceedings of the companion of the 2017 ACM/IEEE international conference on human–robot interaction. ACM, New York, NY, USA, HRI '17, pp 293–294. https://doi.org/10.1145/3029798.3038351
64. Tuffley D (2008) Leadership of complex virtual teams. CSCW 2008 Workshop 9 Supporting Virtual Teamwork
65. Vázquez M, May A, Steinfeld A, Chen WH (2011) A deceptive robot referee in a multiplayer gaming environment. In: 2011 international conference on collaboration technologies and systems (CTS), pp 204 – 211
66. Wang B, Rau PLP (2018) Influence of embodiment and substrate of social robots on users decision-making and attitude. Int J Soc Robot 11:1–11
67. Wijnen L, Coenen J, Grzyb B (2017) “It's not my fault!”: Investigating the effects of the deceptive behaviour of a humanoid robot. In: Proceedings of the companion of the 2017 ACM/IEEE international conference on human–robot interaction. ACM, New York, NY, USA, HRI '17, pp 321–322. https://doi.org/10.1145/3029798.3038300
68. Wu X, Thomas R, Drobina E, Mitzner T, Beer J (2017) An evaluation of a telepresence robot: User testing among older adults with mobility impairment. In: Proceedings of the companion of the 2017 ACM/IEEE international conference on human–robot interaction. ACM, New York, NY, USA, HRI '17, pp 325–326. https://doi.org/10.1145/3029798.3038324
69. Zhou M (2018) Amazon's alexa is going to college. CNET https://www.cnet.com/news/amazons-alexa-is-going-to-college/

Samantha Reig is a PhD student researcher at the Human-Computer Interaction Institute at Carnegie Mellon University, and a NASA Space Technology Research Fellow. She is advised by Jodi Forlizzi and Aaron Steinfeld. Her research interests include socially complex interactions with embodied and disembodied agents, pedestrian interactions with autonomous vehicles, personalized experiences with service robots, and designing for human-robot teaming in space. Prior to beginning her PhD at CMU, Samantha earned a Bachelor of Arts degree in Information Science and Psychology at Cornell University.

Elizabeth Jeanne Carter is a Project Scientist at the Robotics Institute at Carnegie Mellon University who examines how robots and virtual agents can leverage human social skills to improve their interactions with people. Previously, she was a researcher for Disney Research, a division of Walt Disney Imagineering. She has studied interactive avatars, face animation, the uncanny valley, and how the human brain processes social information in children and adults. She received a PhD in Psychology from Carnegie Mellon University.

Xiang Zhi Tan is a Robotics PhD at Carnegie Mellon University, where he is advised by Aaron Steinfeld. He has done research in the areas of prosocial human-robot interaction, assistive robots for blind travelers, and the behavior of multiple robots as people interact with each of them in sequence. Prior to beginning his PhD at CMU, Zhi earned a Bachelor of Science degree in Computer Sciences at the University of Wisconsin–Madison.

Aaron Steinfeld is an Associate Research Professor in the Robotics Institute at Carnegie Mellon University. His research focuses on operator assistance under constraints, i.e., how to enable timely and appropriate interaction when technology use is restricted through design, tasks, the environment, time pressures, and/or user abilities. His work includes intelligent transportation systems, artificial intelligence, human-robot interaction, rehabilitation, and universal design.

Jodi Forlizzi is the Geschke Director and a Professor in the HCI Institute, Carnegie Mellon University, Pittsburgh, PA. Her current research interests include designing services and human-AI collaboration. Jodi's design research has made contributions to social problems including eldercare, accessibility, human assistance, and overall well being.