Robot Errors in Proximate HRI: How Functionality Framing Affects Perceived Reliability and Trust

AURIEL WASHBURN, AKANIMOH ADELEYE, THOMAS AN, and LAUREL D. RIEK,
UC San Diego, San Diego, California

Advancements within human–robot interaction generate increasing opportunities for proximate, goal-directed joint action (GDJA). However, robot errors are common and researchers must determine how to mitigate them. In this article, we examine how expectations for robot functionality affect people’s perceptions of robot reliability and trust for a robot that makes errors. Here 35 participants ($n = 35$) performed a collaborative banner-hanging task with an autonomous mobile manipulator (Toyota HSR). Each participant received either a low- or high-functionality framing for the robot. We then measured how participants perceived the robot’s reliability and trust prior to, during, and after interaction. Functionality framing changed how robot errors affected participant experiences of robot behavior. People with low expectations experienced positive changes in reliability and trust after interacting with the robot, while those with high expectations experienced a negative change in reliability and no change in trust. The low-expectation group also showed greater trust recovery following the robot’s first error compared to the high group. Our findings inform human–robot teaming through: (1) identifying robot presentation factors that can be employed to facilitate trust calibration and (2) establishing the effects of framing, functionality, and the interactions between them to improve dynamic models of human–robot teaming.

CCS Concepts: • **Computer systems organization** → **Robotics**; • **Human-centered computing** → **User studies**; **Collaborative interaction**;

Additional Key Words and Phrases: Human robot teaming, trust, joint action, proximate interaction, robot reliability

ACM Reference format:

Auriel Washburn, Akanimoh Adeleye, Thomas An, and Laurel D. Riek. 2019. Robot Errors in Proximate HRI: How Functionality Framing Affects Perceived Reliability and Trust. *ACM Trans. Hum.-Robot Interact.* 9, 3, Article 19 (May 2020), 21 pages.
<https://doi.org/10.1145/3380783>

1 INTRODUCTION

The rapid development of robotics technology introduces increasing opportunities for humans to work closely with robots across a range of sectors, ranging from housework, to supporting people

This material is based upon work supported by the Air Force Office of Scientific Research (AFOSR) through the award FA9550-18-1-0125, and National Science Foundation (NSF) Graduate Research Fellowship Program under Grant No. DGE-1650112. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or AFOSR.

Authors’ addresses: A. Washburn, A. Adeleye, T. An, and L. D. Riek, UC San Diego, San Diego, California; emails: awashburn@ucdavis.edu, {akadeleye, tan, lriek}@eng.ucsd.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-9522/2019/05-ART19

<https://doi.org/10.1145/3380783>

with disabilities, to aiding people in the workplace, and to helping with clinical procedures. [1, 14, 31, 53, 74]. Across these application domains, people are starting to see robots as teammates instead of tools [47, 63, 72].

Human–robot teams are most successful when there is an appropriate level of trust between the human and the robot [41]. In other words, the level of trust a human places in a robot teammate should be calibrated to the robot’s abilities. Without such trust calibration, people may over-rely on a robot (e.g., neglecting to effectively monitor its behavior), or under-rely on it (e.g., not use it or intentionally disable it). Either of these states can degrade the effectiveness of a team during human–robot interaction (HRI) [18]. HRI researchers agree that supporting trust calibration is central to the future of HRI. However, our field currently has a limited understanding of how trust calibration is established and maintained within a human–robot team [7, 15, 18].

As Hancock et al. [18] note, our ability to achieve better trust calibration is dependent on our understanding of the factors that influence the evolution of trust during interaction. In their meta-analysis of existing work on trust in HRI, they found that robot reliability is one of the primary factors influencing the development of human trust. It is important to note that human–robot trust calibration is not only influenced by robot-related factors, but also by a person’s expectations. These expectations play a large role in a person’s initial experience of a robot [16, 44, 67], critically impacting trust calibration and future collaboration.

Previous work indicates that informing individuals that a robot has limited capabilities can be advantageous to HRI [44]. However, the interaction in this study was open and not goal-directed and did not explore how a person’s expectations might affect their perception of robot capabilities when they are required to engage in collaborative, goal-directed HRI. It seems likely that when an individual’s expectations for robot behavior are not met the individual will see the robot as unreliable and untrustworthy. Ultimately this may interfere with trust calibration and lead to under-reliance on the robot during human-robot teaming.

To our knowledge, there are no studies that examine the influence of functionality framing in the context of proximate, goal-directed HRI. This lack of research is striking, given that a wide variety of rapidly expanding HRI domains, including manufacturing, healthcare, home use, and education, require that human–robot teams engage in proximate, goal-directed joint action (GDJA). Not only is proximate GDJA quickly becoming one of the most common modes of HRI, it is also one of the most challenging from a robotics perspective.

At present, robots are prone to many types of errors [25]. The majority of these errors are due the fact that robots have to operate within dynamic environments using noisy sensors [7]. Proximate GDJA contexts present an especially large number of opportunities for robot errors, as humans themselves contribute to the variability of a teaming environment [8, 22, 26–28, 40, 52].

In this work, we examined the effects of functionality framing on participants’ perceptions of robot reliability and trust during proximate GDJA with a robot exhibiting errors. We split participants into two groups and gave them information about the robot with a low-functionality framing (“the robot is malfunctioning”) or a high-functionality framing (“the robot is fully functional”). All participants experienced a robot making errors during a collaborative task (co-hanging a banner, see Figure 1), regardless of the functionality framing condition they were in. We formulated four central research questions accordingly:

Does functionality framing affect changes in perceived robot reliability before and after interaction with a faulty robot? Given the previously mentioned association between robot reliability and trust, we examined the effects of functionality framing on perceived reliability following robot errors. We expected that before interacting with the robot, individuals in the low-functionality framing group would provide lower reliability ratings than those in the high-functionality framing group. We further predicted that the low-functionality group would display

Fig. 1. The three phases of the task. (Left) The robot and participant start at the marked area and then proceed towards the wall. (Middle) The robot looks toward the participant's standing location to signal it is ready to hang. (Right) The robot then hangs the banner with the participant.

positive changes in perceived reliability following interaction and the high-functionality group would exhibit negative changes.

Does functionality framing affect changes in trust before and after interaction with a faulty robot? We investigated the effects of functionality framing on trust following robot errors in order to improve the current understanding of trust calibration development in GDJA. We expected that the effects of low- vs. high-functionality framing on trust ratings for a robot would be similar to those on perceived reliability (i.e., lower pre-interaction ratings for the low-functionality group, but also a positive change in ratings for the low group and a negative change for the high group).

How does perceived reliability evolve over the course of interaction with a robot making intermittent errors? We aimed to illustrate how robot errors shape perceptions of robot reliability over time to better inform the design of robot presentation and behavior for supporting trust calibration. To do this, we assessed the evolution of perceived reliability over the course of interaction for individuals with low- vs. high-functionality expectations. We predicted that an individual's reports of robot reliability would decrease immediately following robot errors, but that the pattern of recovery following each error might vary depending on the functionality framing they had received.

How does trust evolve over the course of interaction with a robot making intermittent errors? We evaluated the evolution of a person's trust over the course of interaction with the robot in our GDJA task. As for perceived reliability, we expected that robot errors would be followed by decreases in trust, but that the pattern of recovery following each error might be influenced by an individual's functionality framing condition.

Our four research questions were designed to generate an understanding of the relationships between functionality framing, robot errors, perceived reliability, and trust that would allow us to make recommendations for the use of functionality framing to support trust calibration during human-robot teaming. We expected that our functionality framing would have similar effects on both perceived reliability and trust. However, we also predicted that there would be some differences between participants' reports of perceived reliability and trust. To evaluate the convergence and divergence of perceived reliability and trust during our GDJA task, we compared the findings generated by the first and second pairs of questions (i.e., Question 1 and Question 2; Question 3 and Question 4). The results of the first pair of questions allowed us to assess the similarity between changes in trust and perceived reliability for pre- vs. post-interaction questionnaire responses while the outcomes of the second pair of questions provided information about similarities and differences in the evolution of perceived reliability and trust during ongoing interaction.

Our results indicate that functionality framing about robot performance does significantly impact both perceived reliability and trust for a robot co-actor. Individuals with low expectations for

robot functionality exhibited a positive change in both reliability and trust questionnaire responses following interaction.

The contributions of this work are as follows: First, we show that functionality framing impacts how people's expectations in robot behavior align with their experience during interaction through effects on both perceived reliability and trust. Second, we illustrate how functionality framing shapes the actual evolution of human-robot teaming through effects on perceived reliability and trust. This work therefore presents opportunities for transformative advances in proximate HRI by: (1) identifying robot presentation factors that can ultimately be employed to facilitate trust calibration and (2) establishing the effects of framing, functionality, and the interactions between them in order to improve dynamic models of human-robot teaming.

2 BACKGROUND

2.1 Proximate Human-Robot GDJA

Proximate human-robot GDJA includes a wide variety of activities. Sebanz and colleagues [62] suggest that joint action includes all forms of social interaction in which multiple agents coordinate their behaviors in space and time to achieve a desired change in the environment. As a result, the actual physicality of proximate human-robot GDJA spans from shared workspace activities, in which a human and robot are in the same physical space and contribute to a joint goal but don't have direct physical contact with each other or a common object, to co-manipulation tasks in which both the human and robot interact with a single object simultaneously. In between, there are also tasks like handovers, in which the human and robot transfer an object between them.

In addition to this spectrum of physicality among proximate human-robot interaction tasks, the similarity of human and robot roles during interaction can also vary. For example, Peternel and colleagues [46] recently examined human interaction with a robot arm during both collaborative sawing and collaborative surface polishing. In the sawing task, the human and robot alternated the same action to move the saw back and forth; in the surface polishing task, the robot held the weight of the tool while the human controlled the position of the polishing surface.

In our study, a human and robot completed a collaborative co-manipulation task in which they had very similar tasks roles (i.e., each holding one end of a banner and moving toward a wall to hang it on an existing hook). Given this division of labor, the task could not have been completed by either agent alone.

2.2 Robot Expectations

Humans bring many attitudes, beliefs, and previous experiences with technology to their interactions with unfamiliar robots. Together, these factors act to shape the expectations an individual has for a robot prior to interaction. Expectancy Confirmation Theory suggests that expectations consist of the characteristics a person anticipates that a product, service, or technology artifact will have [42, 43]. Expectations are said to directly influence both perceptions of performance and disconfirmation of beliefs for an entity.

As a result, a person's satisfaction with a product, service, or technology artifact following interaction is directly influenced by their perception of its performance and disconfirmation of their prior beliefs. In summary, there is a relationship between pre-adoption expectations and satisfaction that is mediated by the disconfirmation of beliefs.

Robot design also influences people's expectations and consequently affects how they evaluate the robot after interacting with it [9, 12, 48]. This is especially true with respect to the humanness of a robot's appearance.

When the social expectations elicited by humanoid robots are met, people experience interactions with them as enjoyable and empowering [50]. Conversely, users are sometimes disappointed

and dissatisfied when their expectations for human capabilities are not met [10, 38]. This kind of miscalibration between human expectations and robot capabilities based on robot appearance has recently been referred to as the Form Function Attribution Bias (FFAB) [19].

2.3 Framing

User experiences are also shaped by any information they receive prior to HRI about the kind of behaviors a robot is intended to perform [16, 67]. The presentation of such information is either referred to as framing (e.g., [16, 49]), or expectation setting (e.g., [44, 67]). Researchers in psychology suggest that framing or expectation setting helps people to process new information through the activation of contextually relevant knowledge [68, 69]. For example, prior work suggests that giving drones a prevention frame as opposed to a promotion frame leads to increased public support [49]. In that study, people were more supportive of drones if they are presented as “capable of protecting people from harm” vs. “actively seeking out illegal activities to support increased prosecution and punishment”.

Groom et al. [16] suggest that framing in proximate HRI is often as simple as sharing a few facts about a robot prior to interaction. In their work, they found that this process led to greater compliance and confidence in robotic agents during a simulated search and rescue scenario. The authors interpreted this increase in ratings as signifying an increase in trust for framed robots.

Generally positive portrayals of robot abilities have been shown to result in user experiences of dissatisfaction and disappointment for task-oriented robots [44]. In this context, researchers used functionality framing to present a robot as either limited or skilled in a set of abilities. Individuals who expected the robot to be limited exhibited an increase in their perception of robot capabilities following interaction. In contrast, individuals who expected the robot to be skilled displayed a decrease in their perception of robot capabilities after interaction. This outcome of dissatisfaction is predicted by Expectancy Confirmation Theory across contexts where a person’s initial expectations are too high compared to their actual experience [42, 43].

Prior work shows that the expectation for automated aids to perform at nearly perfect rates leads operators to pay too much attention to errors and ultimately underestimate the reliability of the system [11]. This can lead to unwarranted distrust and disuse. Instead, under-representing a robot’s capabilities either through its morphology or the information included in its presentation may provide the greatest potential for successful human–robot teams. In the current study, we evaluate the effect of using framing to transparently or falsely communicate the current functionality of a robot.

2.4 Reliability and Trust

Through a meta-analysis of factors affecting trust in HRI, Hancock et al. [18] demonstrated that performance-based robot characteristics have the largest influence on robot trust. As noted in this meta-analysis, trends in the literature suggest that higher levels of trust are associated with greater robot reliability. Previous work suggests that human trust during HRI is dynamic [45, 58]. This has been especially apparent in the context of interaction with a robot that makes errors [55, 61, 71].

2.5 Robot Errors

Robot errors are common when robots operate in unstructured environments, which include objects and individuals that may frequently change location or behave in other ways that are hard to predict [17, 25, 55, 71]. While it is important to continue making robots more reliable, the complexity of human behavior makes eliminating errors during proximate HRI virtually impossible. The more we know about how robot reliability affects human–robot teaming, the better we will be at achieving effective proximate HRI even when a robot makes errors.

Rovira et al. [59] showed that human–robot team performance on a navigation decision task was already superior to human performance alone even when the robot operated at 80% reliability.

This is consistent with previous work showing that the performance of a human-automation team will be more efficient than a human completing the same task without automation as long as the reliability of the automation is greater than 70% [75]. In cases where 100% reliability cannot always be guaranteed, it is important to explore the effect that robot errors have on user perceptions.

Previous work showed that robot errors significantly impact a human teammate's attributions of robot trust and processes of human-robot trust calibration across HRI [66]. Thus, robot reliability plays a substantial role in shaping people's trust during HRI [18]. Existing work suggests that continuous robot errors can lead to decreases in human trust. However, if robot reliability is high for a period of time at the beginning of an interaction then a robot error may have a less severe effect on trust [13, 32].

Through the use of a real-time trust measure, Desai et al. [6] similarly observed that periods of low robot reliability early on in an interaction had more of an impact on trust than periods of low reliability occurring closer to the middle or end of an interaction. The measurement of trust evolution over the course of an instance of HRI therefore provides valuable information about the impact of robot performance on human experience. Understanding this evolution also allows us to predict future events such as the transfer of trust between HRI tasks [66].

The magnitude of trust decrease following an error, as well as the duration of subsequent trust-recovery, are impacted by the severity of the error along with the frequency of past errors [5]. An individual's perception of error severity is, in turn, likely determined by the task and context. For example, in a goal-oriented task robot errors are especially undesirable and cause users to perceive the robot as less reliable and trustworthy [61]. In time-critical situations, like emergency evacuations, most users completely distrust a robot after a single error [56].

A number of previous studies also indicate that people find robot errors less egregious if the robot acknowledges the error in a way that it is consistent with a typical human response, or is human-like in its appearance or speech [3, 20, 33, 55, 61]. However, because robots do not always know when they have made an error, such acknowledgment is not always possible. There are also situations in which the robot is aware of an error but still cannot explain why the error occurred. Whenever it is possible that a robot will not be able to acknowledge its own errors it is especially critical to provide appropriate framing regarding robot functionality prior to HRI.

2.6 Current Study

In the current study, we compared the effects of pre-interaction framing for low vs. high robot functionality. To do this we evaluated participants' perceived reliability and trust using measures which take into account multiple timescales of experience. As a result, we are able to establish the effect of robot errors on constructs central to human-robot teaming within the context of a GDJA task.

3 METHODOLOGY

In this study, we provided participants with information indicating that the robot's arm was either fully functional or malfunctioning. Then we asked them to hang a banner with the robot (see Figure 1). We programmed the robot to make the same errors each time the experiment was performed, regardless of the participant's framing condition. This experimental design allowed us to establish the effects of an individual's expectations about robot functionality on trust and perceived reliability in the context of robot errors.

3.1 Study Design

For our experiment we employed a Toyota Human Support Robot (HSR). This robot has a mobile base and an arm with five degrees-of-freedom. It stands 1.35 meters tall and has 12 different sensors, including a stereo camera, two wide-angle cameras, a RGB-D sensor, joint angle encoders,

Fig. 2. In our study, a robot and person engaged in proximate, GDJA to collaboratively hang a banner. Intermittently, the robot intentionally dropped the banner 7 cm short of the hook. In these error trials, after dropping the banner the robot always continued to the same final location as in successful trials to give participants' the impression of an unintentional error.

a potentiometer, two force sensors, an IMU, a magnetic sensor, and a LIDAR. The HSR was fully autonomous throughout the current study.

Each participant performed 10 experimental trials. In a single trial, the participant and robot moved toward the wall and hung a 7-foot long banner together. The robot was programmed to hang the banner using predetermined locations and joint configurations. We produced a map of the environment using Hector Mapping [30]. The Toyota HSR's local planner uses this map as a reference frame for goal navigation. At the beginning of an experimental session, we set the robot's initial starting location via RViz. Although we programmed the robot's positions and goal locations ahead of time, the robot did have the ability to sense objects and people around it throughout the experiment. This sensing was primarily used to avoid unforeseen collisions.

Before the start of each trial, a member of the research team placed the banner in the robot's gripper. Once the trial began, the robot moved forward until it reached a preset distance from the wall (Target Location 1). It then lifted its torso, rotated its head to the left where the participant was standing after approaching the wall, and raised its arm. After a brief pause, the base continued to move to a second preset location (Target Location 2). When it reached this location, the robot opened its gripper to release the banner onto the hook.

In trials three and seven, the robot made a pre-programmed error: it dropped the banner on the floor (see Figure 2). To perform the error, the robot exhibited the same behavior as in successful trials until making a premature stop 7 cm short of Target Location 2. It dropped the banner at this position, and then proceeded to Target Location 2 to give the impression of an unintentional error.

We included two error trials, out of the 10 total experimental trials per participant. Thus, the robot's reliability reflected the 80% reliability found to support effective collaborative performance in previous human-automation teams [59]. We introduced the first error in trial three in order to allow participants to build an initial baseline level of trust before the robot made any errors [6]. We placed the second error in trial seven so that we had an equal number of recovery trials following each error (i.e., three) in which to assess the evolution of perceived reliability and trust.

The robot was able to talk through text-to-speech synthesis. It said "starting now" to signify the start of each trial, and "hanging complete" to signify that it had released the banner for hanging. The robot's speech was consistent across all experimental trials. On successful trials, the robot made the "hanging complete" announcement after it had hung its end of the banner the wall. In the error trials, the robot made the announcement after it had dropped the banner on the floor.

The robot did not exhibit any error acknowledgment. As noted earlier, prior work suggests that robot acknowledgment can mitigate the effect of errors [25, 33]; however, in our study, any error

mitigation resulting from robot acknowledgment might have influenced participants' perceptions of robot errors. This would have interfered with our ability to evaluate the effect of functionality framing on the perception of errors. In excluding the possibility of error mitigation in the current study, we are able to provide an initial baseline understanding of the effects of functionality framing on people's experience of robot errors in goal-directed HRI.

3.2 Procedure

When a participant arrived, the experimenter first brought them to a conference room close to the lab. The experimenter then provided the participant with information about the study and the robot. First, the participant received a single page entitled "Human Support Robot (HSR)", which had a picture of the robot and the information that it was an autonomous robot designed to support human activities. On this page we provided three bullet points outlining the robot's ability to: (1) "Sense objects and people around it", (2) "Grab/pick up objects using its gripper arm", and (3) "Move around a room". We confirmed that the participant did not have any questions about this information and then presented a second page, which included the functionality framing information about the robot consistent with the individual's randomly assigned condition. We told participants in the low-functionality condition that the robot's arm was malfunctioning, and those in the high-functionality condition that the robot's arm was fully functional (see Section 3.3).

After the participants completed a manipulation check (see Section 4.1 for details), the experimenter gave them copies of the reliability [29, 65] and trust [39] questionnaires used in the current study. To avoid any influence of the experimenter's presence on participant responses, the experimenter asked the participants to complete the questionnaires alone and meet them in the lab when they were finished.

When the participant entered the lab, the experimenter explained the collaborative GDJA task in view of the robot and experimental test-bed. The banner-hanging task required that the participant and robot each hold one end of the banner, move together across the room to the designated target wall, and hang their respective ends of the banner on existing hooks. The experimenter told the participant that, before each trial began, another member of the research team would hold one end of the banner close to the robot's gripper to initiate robot grasping. The researcher controlled this initial grasping action with a two-key key press via a Bluetooth keyboard. The researcher then handed the other end of the banner to the participant, who always stood to the robot's left. After leaving the test-bed and confirming that the area around the robot was clear of obstacles, the researcher allowed the robot to begin the trial via another two-key press. (After which the robot was fully autonomous.) The participant had an opportunity to do one preliminary practice trial and ask the experimenter any questions they had about the task.

The experimenter then requested that the participant answer the real-time response items addressing perceived reliability and trust after each of the remaining experimental trials. Once the experimenter confirmed that the participant did not have any remaining questions about the task, they explained that the other member of the research team would oversee the preparation of the robot for each trial. The original experimenter then left the room for the duration of the experimental trials. We made this experimental design choice in an effort to (1) reduce the participants' experience of being observed and evaluated during the task and (2) minimize any influence of human-human interaction outcomes on participants' perceptions of robot behavior.

Once the researcher initiated a trial, the robot would say "starting now" before starting to move autonomously toward the target wall. Participants started each trial facing the wall then walked in a straight line towards it parallel to the robot in order to hang the banner. Upon reaching the wall, the robot would lift its torso and arm and then look toward where the participant was standing in front of the left banner hook. We asked the participant to wait until the robot looked toward them

to hang their end of the banner. Once the robot completed its hanging movement sequence for the trial it would say “hanging complete” before turning around and returning to the starting position.

While the robot returned to the starting position, the participant made their real-time reports about whether their perceived reliability and trust for the robot were increasing, decreasing, or remaining the same. Participants made these reports via a sheet of paper on the edge of the test-bed. The robot made the “hanging complete” statement in all trials, including trials three and seven in which it made a pre-programmed dropping error (see Section 3.1 for details). When the robot made the first dropping error, the researcher asked the participants to leave the banner on the floor for them to retrieve.

Once a participant had completed the 10 experimental task trials, the researcher gave them copies of the four-item reliability and trust questionnaires they received before entering the lab for the second time. The researcher then asked the participant to complete the *Willingness to Collaborate* questionnaire. Last, the researcher collected demographic information and the researcher and experimenter debriefed participants about the true nature of the study. A single study session took around 30 to 40 minutes to complete.

We recruited participants through a campus-wide advertisement on a social media group for undergraduate and graduate students. We also sent emails to groups within the Computer Science and Engineering department with the request that information about the study be shared with individuals outside of robotics (i.e., friends, students, staff, and faculty). We made this request in order to minimize the possible biases held by individuals with high levels of robotics familiarity.

In total, 35 individuals participated in the current study. All were fluent English speakers. Seventeen individuals experienced the low-functionality framing condition and 18 experienced the high-functionality framing condition. These sample sizes were chosen as consistent with the majority of existing work in HRI. Participants ranged from 18 to 29 years of age ($M = 26.66$, $SD = 3.00$). Twenty-four participants identified as women and 11 identified as men. Twenty-six participants reported being at least “somewhat familiar” with robotics. All participants provided informed consent. All of the procedures and data collection tools used in the current study were approved by our institutional IRB.

3.3 Variables

Prior to the experiment, participants received information about the robot including details intended to bias their expectations about robot functionality. Participants in the “low” functionality framing condition received a sheet stating that “The robot’s arm is malfunctioning today,” accompanied by a warning symbol. Participants in the “high” functionality framing condition received a sheet stating “The robot’s arm is fully functional today,” accompanied by a check symbol. We also posted signs on the target wall for the banner hanging task reiterating the framing information corresponding to each participant’s condition, consistent with prior HRI work on framing outside of GDJA contexts [44, 67]. We randomly assigned participants’ functionality framing condition within groups of 10 participants such that each group included five individuals in the low condition and five in the high condition.

Before running the current study, we conducted a pilot study on Amazon Mechanical Turk to assess the effects of functionality framing information on perceived reliability and trust. In this pilot, we provided written information about the HSR robot and asked participants to make evaluations based on the hypothetical context of completing a collaborative banner-hanging task with the robot (see Section 3.4 for details on the questionnaire instruments employed). We compared the effects of five different presentation conditions: (1) robot description including mention of errors, (2) robot description including mention of property damage, (3) robot description including mention of both errors and property damage, (4) robot description including mention of errors and

Table 1. Trust and Reliability Questionnaires

Trust Questionnaire

1. I can depend on the robot to work correctly every time.
2. The robot seems reliable.
3. If I did the same task with the robot again it would do it the same way.
4. I could trust the robot to work whenever I might need it.

Reliability Questionnaire

1. To what extent can the robot's behavior be predicted from moment to moment?
2. To what extent can you count on the robot to do its job?
3. What degree of faith do you have that the robot will be able to cope with similar situations in the future?
4. Overall how much do you trust the robot?

property damage followed by an additional page stating that the robot was malfunctioning that day, and (5) robot description without any mention of errors or property damage followed by an additional page stating that the robot was malfunctioning that day.

Results from a total of 41 participants, who were each randomly assigned to one of the five conditions, demonstrated that perceived reliability and trust were lowest for condition 5. This indicated that a poster about the current functionality status of the robot was the most effective way to frame robot shortcomings so that participants would expect reduced reliability and trustworthiness during interaction. These findings therefore informed the effectiveness of the framing language and presentation used in the current study.

3.4 Measurements

We used validated questionnaires to assess perceived robot reliability [29, 65] and trust in the robot [39] before and after interaction with the HSR. We also obtained real-time measures of reliability and trust following each experimental trial.

Each validated questionnaire consisted of four items (see Table 1). We asked participants to respond to these questionnaires before they interacted with the robot (i.e., directly after the manipulation check, which is presented in Section 4.1) and again after they completed all of the interactive experimental trials. Before interaction with the robot, we prefaced each questionnaire with the statement: "Imagine you're going to hang a banner with this robot. Based on the information you just learned, please respond to the following questions." After participants completed the banner-hanging trials, we introduced each questionnaire with the language "Now that you have finished performing the task with the robot, please update your ratings."

The reliability questionnaire consisted of four items [29, 65]. We provided response options to these items as discrete visual-analogue scales (DVASs) with six options. The scales ranged from "Strongly Disagree" to "Strongly Agree", with only the endpoints labelled. We also used a trust questionnaire made up of four items [39]. We also provided DVASs for these items with response options ranging from "Not at all" to "Completely" with only the endpoints labeled. We evaluated each questionnaire by averaging the numeric values corresponding to an individual's response selections for each of the four items. High scores indicated greater reliability and trust on the respective questionnaire instruments.

Using the real-time measures of reliability and trust administered following each experimental trial, we generated an event-contingent series for the evolution of each construct over the course of interaction. These series are similar to those obtained through Ecological Momentary Assessment (EMA) paradigms [64]. For these real-time measurements, we had participants indicate whether

their reliability and trust in the robot were increasing ($\uparrow$), decreasing ($\downarrow$), or remaining the same ($\leftrightarrow$) after each iteration of the banner-hanging task. We adapted this response measure from one designed by Desai et al. [6]. We administered it here via paper with reliability presented on the top half of a single sheet of paper, and trust presented on the bottom half. There was also a place for participants to write comments associated with each rating.

It is important to note that we were interested in generating information about people's perceived reliability and trust for the robot based on their own internal conceptions of each construct. In order to do this, we did not define "reliability" and "trust" for participants at any point during the study.

After participants completed the post-interaction reliability and trust questionnaires, we asked them to complete a questionnaire on their willingness to collaborate with the robot on a number of additional tasks (see [Appendix](#)). We created this questionnaire based on tasks Beer et al. [1] had previously identified as tasks for which older adults would prefer robot assistance to human assistance. This consisted of 14 home tasks, 10 of which are primarily physical (cleaning a kitchen, cleaning a bathroom, cleaning a window, pest control, loading dishes in a dishwasher, making the bed, cleaning the floor, taking out the trash, doing laundry) and four of which are solely cognitive (reminders to take medication, staying informed on weather/news, learning how to use technology, getting information about your hobbies). For each task, participants indicated their willingness to collaborate with the robot using a six-option DVAS, ranging from "Not at all" to "Completely", with only the endpoints labeled. While we recognize that this measure has not undergone a reliability analysis, we thought that it could still provide valuable information in the context of the current study.

4 RESULTS

4.1 Manipulation Check

Each participant in the current study was asked to complete a manipulation check regarding their functionality framing condition before they entered the lab. After we presented the introductory information about the robot, we asked each participant to respond to the questions "Is the robot fully functional today?" and "If no, what's wrong with it?". All participants correctly answered the manipulation check questions based on their framing condition.

4.2 Pre- vs. Post-interaction Reliability and Trust

We used the four-item reliability questionnaire of Sidner et al. [65] and the four-item trust scale of Muir [39] to assess participants' perceived reliability and trust both before and after completing all 10 experimental banner-hanging trials with the robot. Based on a participant's response to each of these measures, we calculated their average pre- and post-interaction reliability and trust scores. We then used these pre- and post-interaction scores to calculate the change in an individual's reliability and trust following interaction by subtracting their pre-interaction scores from their post-interaction scores. We averaged these change scores by functionality framing group so that we could compare the effect of functionality framing on changes in perceived reliability and trust following interaction.

A paired t-test revealed a significant effect of functionality framing on reliability change scores, $t(33) = 3.58$, $p = .001$, *Cohen's d* = 1.25. Here, participants in the low-functionality condition reported an increase in the reliability of the robot following interaction while those in the high functionality condition reported a decrease (see [Figure 3\(A\)](#)).

A paired t-test on trust change scores also revealed a significant effect of functionality framing, $t(33) = 3.32$, $p = .002$, *Cohen's d* = 1.15. Participants in the low-functionality condition reported

Fig. 3. Average change in reliability (A) and trust (B) for post vs. pre-interaction in the low- and high-functionality framing groups. Error bars show standard error. $**p < .01$.

an increase in trust following interaction while participants in the high functionality condition reported no change in trust (see Figure 3(B)).

4.3 Real-Time Reliability and Trust

We also asked participants to provide real-time, event-contingent reports of reliability and trust during interaction with the robot. Participants completed these assessments after each experimental trial, indicating whether the reliability of the robot and their trust in the robot were increasing, decreasing, or remaining the same [6]. We initially assigned directional values to each response category, using +1 for “increasing”, -1 for “decreasing”, and 0 for “remaining the same”. We then used these directional values to estimate separate construct evolution series for reliability and trust as they changed over the course of an individual’s interaction with the robot. The first value of each construct evolution series was the directional value corresponding to a participant’s response for that construct (i.e., perceived reliability or trust) after the first experimental trial. We then calculated all subsequent values in the series by finding the sum of the directional value for the current trial and the construct evolution value for the previous trial (e.g., if the participant’s trust was “remaining the same” after the first trial and “increasing” after the second trial, then the first two values of the construct evolution series for trust would be 0 and 1).

The evolution of real-time perceived reliability and trust can be seen in Figures 4(A, C) (respectively). We statistically evaluated the effect of framing condition on each of perceived reliability and trust evolution by treating each error trial as an anchoring event [64] and evaluating changes in the three trials following that error. To do this, we calculated the normalized area under the curve (AUTC) [6] for each construct over each time period of interest (i.e., trials 4-6 for Error 1 and trials 8-10 for Error 2) using the trapezoidal rule (see [21]). We then used an Analysis of Variance (ANOVA) to compare the AUTC for each framing group at each of the timepoints, as is common practice within the analysis of EMA data [2].

A 2 (functionality framing: low, high) $\times$ 2 (robot error: 1, 2) mixed model ANOVA on the AUTC for perceived reliability evaluation revealed no significant interaction between the two variables or main effects (see Figure 4(B)). However, a 2 (functionality framing: low, high) $\times$ 2 (robot error: 1, 2) mixed model ANOVA on the AUTC for trust evolution did reveal a significant interaction between framing and robot error, $F(1, 33) = 5.70$, $p = .02$, $\eta p^2 = .15$ (see Figure 4(D)). Pairwise comparisons confirmed that following Error 1 trust increased significantly more in the low- vs. high-functionality group ($p = .045$). Pairwise comparisons also showed that, within the low-functionality group, there was a significantly greater increase in trust following the Error 1 compared to Error 2 ($p = .01$). These results demonstrate that the interaction between functionality

Fig. 4. Evolution of reliability (A) and trust (C) during interaction for the low and high expectation groups. Dashed lines show trials in which the robot made a banner-dropping error (Trials 3 and 7). Average reliability (B) and trust (D) evolution following each of the two robot errors for participants in the low and high functionality framing conditions. Error bars show standard error.

framing and robot error number was driven by the fact that participants in the low-functionality group showed a greater increase in trust compared to the high-functionality group after Error 1, but not after Error 2. Additionally, participants in the low-functionality group showed a greater increase in trust following Error 1 compared to Error 2, while the high-functionality group did not show significantly different levels of trust increase following Error 1 vs. Error 2.

4.4 Collaboration Transfer

After participants completed all ten trials of the banner-hanging task and the post-interaction reliability and trust questionnaires, they responded to our 14-item *Willingness to Collaborate* questionnaire (see Section 3.4 for details). For each participant, we generated a summed measure for each of the two task domains represented in the questionnaire (i.e., physical vs. cognitive), with a maximum possible sum of 60 for physical tasks and 24 for cognitive tasks. We then divided each participant's sum in each task domain by these possible maxima to obtain their percent willingness to interact in the future. We averaged this percent willingness across participants in each framing condition. This allowed us to conduct paired t-tests comparing the framing conditions within each domain.

There was no significant effect of framing group on willingness to collaborate on either physical ($t(33) = .22, p = .83$) or cognitive tasks ($t(33) = 1.5, p = .14$). The average willingness to collaborate on additional physical tasks was 73.04% for the low-functionality group (SD = 3.91%) and 74.19% for the high-functionality group (SD = 3.58%). The average willingness to collaborate on

additional cognitive tasks was 70.97% for the low-functionality group ($SD = 5.96\%$) and 82.18% for the high-functionality group ($SD = 4.51\%$).

5 DISCUSSION

HRI researchers, industry professionals, and consumers are all eager to have robots that can engage in proximate HRI within human environments. Unfortunately, human behavior is largely unpredictable, and these environments change frequently and rapidly. These conditions are challenging for robots, leading to problems ranging from sensor occlusion to unsuccessful behavioral planning [25]. As a result, robot errors are common in proximate human-robot joint action. While it is important to continue improving robot resiliency to the challenges of proximate HRI, it is also critical that we learn how to support effective human-robot teaming in the context of robot errors.

In our study, we experimentally manipulated individuals' expectations for the functionality of a robot. We then addressed four specific research questions about the effects of a person's expectations on their perceptions of a robot's reliability and trustworthiness after the robot made errors. Our first research question was whether functionality framing would affect changes in perceived reliability following collaborative interaction with the robot. Our findings revealed that individuals with low expectations thought the robot was more reliable after they interacted with it than they did before they interacted. In contrast, those with high expectations thought the robot was less reliable after they interacted with it.

Our second research question was whether functionality framing would affect changes in trust following collaboration with a robot making errors. Here we also found that the change in participants' post- vs. pre-interaction questionnaire responses was affected by functionality framing. Individuals in the low-functionality framing group showed an increase in trust ratings following interaction with the robot, while individuals in the high-functionality framing group showed no change in trust.

The collective findings from our first two research questions showed that individuals in the low-functionality group displayed positive changes in both perceived reliability and trust following interaction while individuals in the high-functionality group showed a negative change in reliability and no change in trust. These results are consistent with previous empirical work showing that human-robot trust is strongly associated with robot reliability [18]. However, our findings also support the idea that trust is robust to some amount of change in robot reliability over time [6]. We expect that the relationship between robot reliability and trust during HRI is largely shaped by the nature of the task. For example, the participants in our study reported that the robot's reliability rate was acceptable for the banner-hanging task, but would not have been acceptable for a more delicate task. They communicated this with statements like the following, which came from three unique participants:

"The robot is overall reliable, since it only failed twice. However, I wouldn't trust it with delicate objects, such as glass, unless it makes zero errors."

"The robot could be used for simple/non-delicate tasks just fine with reliability as it is."

"It would be harder to trust the robot if there was a delicate object involved."

In response to research questions (1) and (2), we predicted that individuals with low expectations for robot functionality would experience more positive changes in their perceptions of the robot following interaction compared to individuals with high-functionality expectations. This prediction was confirmed by individuals' pre- vs. post-interaction questionnaire responses for perceived

reliability and trust. We suggest that anyone in the position of presenting a robot for proximate, goal-directed HRI to new users should be careful not to overstate the robot's capabilities, as this can lead to negative changes in perceived reliability and trust during teaming.

Relatedly, since robot humanness appears to induce high expectations for functionality [10, 38] we recommend that designers and roboticists consider robot morphology very carefully. In general, both the design and presentation of a robot should take into account the detrimental effects a miscalibration between human expectations and robot capabilities can have on human-robot team dynamics. This idea is consistent with our prior work suggesting that the use of humanoid portrayals shapes human expectations in a way that could constitute manipulation, and that the associated ethical implications should be taken seriously [51, 54].

In our third and fourth research questions, we asked whether functionality framing would influence the evolution of perceived reliability and trust during interaction. To address these questions, we asked participants to make real-time reports about their perception of the robot's reliability and their trust in the robot after each iteration of the banner-hanging task. These reports revealed that participants' perceived reliability and trust for the robot increased across trials in which the robot performed the task successfully, and decreased immediately following each of the two pre-programmed errors in robot behavior. This trajectory was similar across framing groups.

When we evaluated our third research question, we saw no effect of framing group on the evolution of individuals' perceived robot reliability. This suggests that functionality framing may not affect an individual's perception of robot reliability on a trial-by-trial basis. However, when we evaluated our fourth research question we observed an interaction between functionality framing and trust. Individuals in the low-functionality framing group reported a greater increase in trust following the first robot error as compared to the second robot error. In contrast, individuals in the high-functionality framing group reported similar trust increases following the first and second error.

These findings indicate that individuals with low expectations exhibit greater trust recovery following an initial robot error compared to those with high expectations. Previous work in robot tele-operation has revealed that a user's trust typically increases faster during an initial stable period of high robot reliability than it does after a drop in robot reliability [6]. Our own results suggest that low expectations may mitigate the negative effect that a single robot error has on trust building. Despite our observation that functionality framing has significant effects on changes in perceived reliability and trust over the course of interaction, functionality framing did not appear to have an effect on the willingness of individuals to collaborate with the same robot on future tasks.

5.1 Implications across Proximate GDJA Tasks

Notably, proximate GDJA is not only shaped by task physicality and co-actor roles, but also by the level of robot responsiveness to changes in task conditions, including those arising from human actions. The current task did not require robot adaptation to environmental changes or human actions beyond the ability to avoid collisions. As a result, the robot's level of responsiveness was relatively basic.

While many human-robot interaction tasks will benefit from increased robot responsiveness, robots still have limited autonomy and require high levels of assistance from human collaborators in a wide range of current joint human-robot tasks. For example, robot arms used in automobile manufacturing have improved human-robot interaction by enabling robots to avoid collisions and increase human safety, but human collaborators often program the robots to carry out very specific, repetitive tasks and must ensure that the materials needed for a task are located in consistent, designated locations. In fact, in our own recent work, we highlight the fact that

technical limitations surrounding effective robot responsiveness in dynamic human environments is currently one of the biggest challenges to the advancement of human-robot teaming [35, 36, 73]. Accordingly, we and other HRI researchers are now working on new methods to build more responsive robots. Because of the difficulty associated with achieving successful robot responsiveness in HRI contexts, previous researchers investigating human perceptions of robots and trust in human-robot teaming have primarily used methods for which robot responsiveness to human behaviors is low. Many of these studies did not even include physical robots, but instead rely on videos of robot behavior (e.g., [3]), or simulated or virtual robots (e.g., [55, 56, 71]). Others involve manual control over all robot behavior or robot responses to human behavior via Wizard-of-Oz (e.g., [57, 61]). While the HRI community should continue to prioritize the development of more responsive robots for human-robot teaming, our use of an autonomous physical, albeit minimally responsive, robot engaged in co-manipulation is relatively advanced within the study of human perception of robots and trust in human-robot teaming. The development and evolution of teaming dynamics like trust and reliability are likely to be influenced by the level of robot responsiveness during interaction. In this study we demonstrate that for a collaborative task in which a low level of robot responsiveness is sufficient to achieve the joint goal, human expectations for robot functionality shape how these teaming dynamics are affected by robot errors.

We expect that, for tasks in which the robot must be more responsive to human actions, robot errors will have an even greater impact on perceived reliability and trust. Similarly, we predict that robot framing and presentation will be even more critical to supporting effective teaming in these contexts by affording appropriate expectation setting for robot abilities and behaviors.

5.2 Participant Speculations about Error Causation

Interestingly, several participants across both functionality framing conditions reported that after the robot made an error they began to closely analyze its behavior. Participants wanted to understand why an error had occurred. They also sometimes thought they could predict when errors were going to occur. Many participants incorrectly identified factors they thought were associated with the robot dropping errors, but some did accurately note a change in robot arm behavior associated with robot dropping. Specifically, if the robot raised its arm and then moved forward and stopped short of the normal location where it released the banner onto the hook (i.e., Target Location 2) it was likely to immediately drop the banner on the floor. Ultimately, this change in robot behavior associated with the dropping error could not be detected until just a few seconds beforehand.

Still, some individuals attributed robot errors to their own behavior early on in a trial (e.g., walking too fast with the banner before reach the robot's first stopping location). This feeling of responsibility for robot errors may be due to the fact that the human and robot had very similar roles in the banner-hanging task. Past work suggests that people feel more responsible for a task when working with a peer or subordinate robot than when working with a supervisor robot [23]. As a result, people may be more likely to assume they have an effect on robot errors in tasks where human and robot roles are very similar compared to those in which the robot has a more subordinate role. Ultimately, the participant feedback we received in the current study emphasized that providing information about the cause of robot errors during HRI may reduce the attentional resources a human co-actor uses in response to robot errors.

5.3 Limitations and Future Directions

Our participants did not experience any acute risk during interaction with the robot. This may limit our ability to contribute to a common discussion around risk and trust in HRI. Many models of trust within HRI include risk as an important component in trust development. These models

are often based on a definition of trust offered by Mayer and colleagues [37]. In this definition, trust is based on a person's willingness to be vulnerable to the actions of another party irrespective of the ability to control that party's actions.

The current study requires that individuals engage in human-robot teaming in order to complete a GDJA HRI task. Human-robot teaming itself requires a human to depend on a robot to complete a portion of the collective task. More specifically, the human cannot control the robot's actions and the team's success is vulnerable to the robot's behavior [34]. As a result, our participants experienced a low level of vulnerability during interaction, but no explicit risk. Future studies of human-robot trust during GDJA will benefit from the thoughtful introduction of additional risk to participants.

In order to identify how functionality framing information might impact human-robot team performance in the future we assessed participants' willingness to collaborate with the robot on additional tasks following the completion of the banner-hanging experiment. It is possible that our ability to capture the effect of functionality framing on future collaboration was limited by our use of a self-report measure. Previous work suggests that self-report statements do not always fully represent how an individual would actually act in a given situation [4]. Alternatively, implicit, behavioral metrics can allow researchers to measure constructs, such as trust, frequently and unobtrusively during ongoing interaction (e.g., see Hayes et al. [20]). In future work, researchers may be able identify a relationship between functionality framing and willingness to collaborate using an implicit measure of willingness to collaborate.

It is also possible that we observed similar responses between framing groups on our willingness to collaborate questionnaire because interaction experience is more important than functionality framing in determining a person's willingness to collaborate on future tasks. In other words, participants may have made decisions about collaborating with the robot in the future based on what they learned about the robot's capabilities during interaction instead of the framing information they received about the robot at the beginning of the experiment.

Interestingly, our participants spent considerable time thinking about potential causes for the robot's errors. We chose not to include error acknowledgment in this study as this might have mitigated the effect of robot errors on participants' perceptions of reliability and trust. This allowed us to provide a baseline illustration of the relationship between functionality framing and error during proximate HRI. However, previous work suggests that robot transparency surrounding errors is beneficial to human experience and team performance during HRI [6, 24, 60, 70] and feedback from our own participants indicates that this would be valuable within proximate, GDJA as well. Thus, future work on human-robot teaming fluency in proximate, GDJA should aim to establish how robot responses can reduce the attentional resources humans give to robot errors when they do occur.

Last, in our current study we used just two contrasting presentation conditions to evaluate the effects of functionality framing on people's experiences of an error-prone robot. HRI researchers should continue to empirically evaluate other presentation styles, including those in which presenters do not provide any explicit information about the functionality of a robot, to gain a better understanding of the effects of robot presentation on the perceived reliability and trustworthiness of robots following errors.

5.4 Conclusion

Our results suggest that in HRI contexts where robot errors are probable human co-actors are likely to be disappointed by robot capabilities unless they are told ahead of time that the robot may make errors. This effect likely impacts human-robot teaming in a wide range of proximate GDJA tasks within manufacturing, healthcare, home use, and educational contexts. We also showed that

when people expect robot errors they exhibit greater trust recovery after a single robot error. Our measures revealed a strong relationship between trust and reliability, while also indicating that there is independence between the two constructs.

Participants in the current study tended to generate explanations about why the robot made errors and to try to predict when the next error would occur. This observation supports the idea that future research should continue to examine how robots can acknowledge errors during interaction in order to minimize human attention to them. As we have noted, future work should also systematically explore the effect of risk on reliability and trust in proximate GDJA. In conjunction with our current work on robot presentation, these avenues will greatly expand our ability to support trust calibration across a range of proximate HRI contexts in which robot errors are known to be frequent and impactful events. At present, our findings can be utilized for the continued improvement of (1) framing factors for robots to be used in proximate HRI and (2) modeling of human–robot trust for the support of effective human–robot teaming.

A APPENDIX

Willingness to Collaborate Questionnaire

Physical Tasks

1. Do you trust the robot to help you change a light bulb?
2. Do you trust the robot to help you clean a kitchen?
3. Do you trust the robot to help you clean a bathroom?
4. Do you trust the robot to help you clean a window?
5. Do you trust the robot to help you with pest control?
6. Do you trust the robot to help you load dishes in a dishwasher?
7. Do you trust the robot to help you make a bed?
8. Do you trust the robot to help you clean the floor?
9. Do you trust the robot to help you take out the trash?
10. Do you trust the robot to help you do laundry?

Cognitive Tasks

11. Do you trust the robot to remind you to take medication?
 12. Do you trust the robot to help you stay informed about the weather/news?
 13. Do you trust the robot to help you learn how to use new technology?
 14. Do you trust the robot to help you get information about your hobbies?
-

REFERENCES

- [1] Jenay M. Beer, Cory-Ann Smarr, Tiffany L. Chen, Akanksha Prakash, Tracy L. Mitzner, Charles C. Kemp, and Wendy A. Rogers. 2012. The domesticated robot: Design guidelines for assisting older adults to age in place. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 335–342.
- [2] Elin A. Björling and Narayan Singh. 2017. Exploring temporal patterns of stress in adolescent girls with headache. *Stress and Health* 33, 1 (2017), 69–79.
- [3] Elizabeth Cha, Anca D. Dragan, and Siddhartha S. Srinivasa. 2015. Perceived robot capability. In *Proceedings of the 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 541–548.
- [4] J. Y. C. Chen and P. I. Terrence. 2009. Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics* 52, 8 (2009), 907–920.
- [5] Cynthia L. Corritore, Beverly Kracher, and Susan Wiedenbeck. 2003. On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies* 58, 6 (2003), 737–758.
- [6] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. 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, 251–258.

- [7] Munjal Desai, Kristen Stubbs, Aaron Steinfeld, and Holly Yanco. 2009. Creating trustworthy robots: Lessons and inspirations from automated systems. In *Proceedings of the AISB Convention: New Frontiers in Human-robot Interaction*. 6–9.
- [8] Sandra Devin and Rachid Alami. 2016. An implemented theory of mind to improve human-robot shared plans execution. In *Proceedings of the 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 319–326.
- [9] Carl F. DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. 2002. All robots are not created equal: The design and perception of humanoid robot heads. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. ACM, 321–326.
- [10] Brian R. Duffy. 2003. Anthropomorphism and the social robot. *Robotics and Autonomous Systems* 42, 3-4 (2003), 177–190.
- [11] Mary T. Dzindolet, Linda G. Pierce, Hall P. Beck, Lloyd A. Dawe, and B. Wayne Anderson. 2001. Predicting misuse and disuse of combat identification systems. *Military Psychology* 13, 3 (2001), 147.
- [12] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3–4 (2003), 143–166.
- [13] Jean Fox and Deborah Boehm-Davis. 1998. Effects of age and congestion information accuracy of advanced traveler information systems on user trust and compliance. *Transportation Research Record: Journal of the Transportation Research Board* 1621 (1998), 43–49.
- [14] Matthew Gombolay, Xi Jessie Yang, Bradley Hayes, Nicole Seo, Zixi Liu, Samir Wadhwania, Tania Yu, Neel Shah, Toni Golen, and Julie Shah. 2018. Robotic assistance in the coordination of patient care. *The International Journal of Robotics Research* 37, 10 (2018), 1300–1316.
- [15] Victoria Groom and Clifford Nass. 2007. Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies* 8, 3 (2007), 483–500.
- [16] Victoria Groom, Vasant Srinivasan, Cindy L. Bethel, Robin Murphy, Lorin Dole, and Clifford Nass. 2011. Responses to robot social roles and social role framing. In *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 194–203.
- [17] Adriana Hamacher, Nadia Bianchi-Berthouze, Anthony G. Pipe, and Kerstin Eder. 2016. Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction. In *Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 493–500.
- [18] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (2011), 517–527.
- [19] Kerstin S. Haring, Katsumi Watanabe, Mari Velonaki, Chad C. Tossell, and Victor Finomore. 2018. FFAB-The form function attribution bias in human-robot interaction. *IEEE Transactions on Cognitive and Developmental Systems* (2018).
- [20] Cory J. Hayes, Maryam Moosaei, and Laurel D. Riek. 2016. Exploring implicit human responses to robot mistakes in a learning from demonstration task. In *Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 246–252.
- [21] P. M. Herbers, D. A. Elder, and J. G. Woo. 2011. Area under a curve: Calculation and visualization. In *Proceedings of the 2011 Midwest SAS Users Group. (Report No.: DG12a2011)*.
- [22] Laura M. Hiatt, Anthony M. Harrison, and J. Gregory Trafton. 2011. Accommodating human variability in human-robot teams through theory of mind. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*.
- [23] Pamela J. Hinds, Teresa L. Roberts, and Hank Jones. 2004. Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human-Computer Interaction* 19, 1 (2004), 151–181.
- [24] Guy Hoffman and Cynthia Breazeal. 2008. Anticipatory perceptual simulation for human-robot joint practice: Theory and application study. In *Proceedings of AAAI* 1357–1362.
- [25] Shanee Sarah Honig and Tal Oron-Gilad. 2018. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in Psychology* 9 (2018), 861.
- [26] Tariq Iqbal, Michael J. Gonzales, and Laurel D. Riek. 2015. Joint action perception to enable fluent human-robot teamwork. In *Proceedings of the 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 400–406.
- [27] Tariq Iqbal, Samantha Rack, and Laurel D. Riek. 2016. Movement coordination in human–robot teams: A dynamical systems approach. *IEEE Transactions on Robotics* 32, 4 (2016), 909–919.
- [28] T. Iqbal and L. D. Riek. 2017. Human robot teaming: Approaches from joint action and dynamical systems. *Humanoid Robotics: A Reference* (2017).
- [29] Cory David Kidd. 2003. *Sociable Robots: The Role of Presence and Task in Human-Robot Interaction*. Ph.D. Dissertation. Massachusetts Institute of Technology.

- [30] Stefan Kohlbrecher, Johannes Meyer, Thorsten Graber, Karen Petersen, Uwe Klingauf, and Oskar von Stryk. 2013. Hector open source modules for autonomous mapping and navigation with rescue robots. In *Robot Soccer World Cup*. Springer, 624–631.
- [31] Anthony R. Lanfranco, Andres E. Castellanos, Jaydev P. Desai, and William C. Meyers. 2004. Robotic surgery: A current perspective. *Annals of Surgery* 239, 1 (2004), 14.
- [32] John D. Lee and Katrina A. See. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors* 46, 1 (2004), 50–80.
- [33] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully mitigating breakdowns in robotic services. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 203–210.
- [34] Michael Lewis, Katia Sycara, and Phillip Walker. 2018. The role of trust in human-robot interaction. In *Foundations of Trusted Autonomy*, Husein A. Abbass, Jason Scholz, and Darryn J. Reid (Eds.). Springer-International, Cham, Switzerland, 135–159.
- [35] Sachiko Matsumoto and Laurel D. Riek. 2019. Fluent coordination in proximate human robot teaming. In *AI and Its Alternatives for Shared Autonomy in Assistive and Collaborative Robotics, Workshop at Robotics: Science and Systems (RSS)*.
- [36] Sachiko Matsumoto, Auriel Washburn, and Laurel D. Riek. in review. A framework to explore proximate human-robot coordination.
- [37] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An integrative model of organizational trust. *Academy of Management Review* 20, 3 (1995), 709–734.
- [38] Masahiro Mori. 1970. The uncanny valley. *Energy* 7, 4 (1970), 33–35.
- [39] Bonnie M. Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (1994), 1905–1922.
- [40] Aastha Nigam and Laurel D. Riek. 2015. Social context perception for mobile robots. In *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3621–3627.
- [41] Sherry Ogretan, Stephanie Lackey, and Denise Nicholson. 2010. Recommended roles for uninhabited team members within mixed-initiative combat teams. In *Proceedings of the 2010 International Symposium on Collaborative Technologies and Systems (CTS)*. IEEE, 531–536.
- [42] Richard L. Oliver. 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research* (1980), 460–469.
- [43] Richard L. Oliver and Gerald Linda. 1981. Effect of satisfaction and its antecedents on consumer preference and intention. *ACR North American Advances* (1981).
- [44] Steffi Paepcke and Leila Takayama. 2010. Judging a bot by its cover: An experiment on expectation setting for personal robots. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*. IEEE Press, 45–52.
- [45] Eui Park, Quaneisha Jenkins, and Xiaochun Jiang. 2008. Measuring trust of human operators in new generation rescue robots. In *Proceedings of the JFPS International Symposium on Fluid Power* 2008, 7–2 (2008), 489–492.
- [46] Luka Peternel, Nikos Tzagarakis, Darwin Caldwell, and Arash Ajoudani. 2018. Robot adaptation to human physical fatigue in human–robot co-manipulation. *Autonomous Robots* 42, 5 (2018), 1011–1021.
- [47] Elizabeth Phillips, Scott Ososky, Janna Grove, and Florian Jentsch. 2011. From tools to teammates: Toward the development of appropriate mental models for intelligent robots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55. SAGE Publications, Sage, CA: Los Angeles, CA, 1491–1495.
- [48] Aaron Powers and Sara Kiesler. 2006. The advisor robot: Tracing people’s mental model from a robot’s physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*. ACM, 218–225.
- [49] Lisa M. Pytlik-Zillig, Brittany Duncan, Sebastian Elbaum, and Carrick Detweiler. 2018. A drone by any other name: Purposes, end-user trustworthiness, and framing, but not terminology, affect public support for drones. *IEEE Technology and Society Magazine* 37, 1 (2018), 80–91.
- [50] Byron Reeves and Clifford Ivar Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press.
- [51] Laurel Riek and Don Howard. 2014. A code of ethics for the human-robot interaction profession. In *Proceedings of We Robot* (2014).
- [52] Laurel D. Riek. 2013. The social co-robotics problem space: Six key challenges. In *Proceedings of the Symposium on Robotics Challenges and Vision (RCV’13)* (2013).
- [53] Laurel D. Riek. 2017. Healthcare robotics. *Commun. ACM* 60, 11 (2017), 68–78.
- [54] Laurel D. Riek, Woodrow Hartzog, Don A. Howard, AJung Moon, and Ryan Calo. 2015. The emerging policy and ethics of human robot interaction. In *HRI (Extended Abstracts)*. 247–248.
- [55] Paul Robinette, Ayanna M. Howard, and Alan R. Wagner. 2015. Timing is key for robot trust repair. In *Proceedings of the International Conference on Social Robotics*. Springer, 574–583.

- [56] Paul Robinette, Ayanna M. Howard, and Alan R. Wagner. 2017. Effect of robot performance on human–robot trust in time-critical situations. *IEEE Transactions on Human-Machine Systems* 47, 4 (2017), 425–436.
- [57] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *Proceedings of the 11th ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 101–108.
- [58] Denise M. Rousseau, Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. Not so different after all: A cross-discipline view of trust. *Academy of Management Review* 23, 3 (1998), 393–404.
- [59] Ericka Rovira, Austin Cross, Evan Leitch, and Craig Bonaceto. 2014. Displaying contextual information reduces the costs of imperfect decision automation in rapid retasking of ISR assets. *Human Factors* 56, 6 (2014), 1036–1049.
- [60] Dorsa Sadigh, Shankar Sastry, Sanjit A. Seshia, and Anca D. Dragan. 2016. Planning for autonomous cars that leverage effects on human actions. In *Robotics: Science and Systems*.
- [61] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 141–148.
- [62] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. 2006. Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences* 10, 2 (2006), 70–76.
- [63] Kerstin Severinsson-Eklundh, Anders Green, and Helge Hüttenrauch. 2003. Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous Systems* 42, 3–4 (2003), 223–234.
- [64] Saul Shiffman. 2013. Conceptualizing analyses of ecological momentary assessment data. *Nicotine & Tobacco Research* 16, Suppl_2 (2013), S76–S87.
- [65] Candace L. Sidner, Christopher Lee, Cory D. Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166, 1–2 (2005), 140–164.
- [66] Harold Soh, Pan Shu, Min Chen, and David Hsu. 2018. The transfer of human trust in robot capabilities across tasks. *Arxiv Preprint Arxiv:1807.01866* (2018).
- [67] Yuan Sun. 2011. *Effects of Expectation Setting and Agency on Human-Robot Interaction*. Master’s thesis. The Pennsylvania State University, State College, Pennsylvania.
- [68] Deborah Tannen. 1993. What’s in a frame? Surface evidence for underlying expectations. *Framing in Discourse* 14 (1993), 56.
- [69] Deborah Tannen and Cynthia Wallat. 1987. Interactive frames and knowledge schemas in interaction: Examples from a medical examination/interview. *Social Psychology Quarterly* (1987), 205–216.
- [70] Ning Wang, David V. Pynadath, and Susan G. Hill. 2016. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *Proceedings of the 11th ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 109–116.
- [71] Ning Wang, David V. Pynadath, Ericka Rovira, Michael J. Barnes, and Susan G. Hill. 2018. Is it my looks? or something I said? The impact of explanations, embodiment, and expectations on trust and performance in human-robot teams. In *Proceedings of the International Conference on Persuasive Technology*. Springer, 56–69.
- [72] Samantha F. Warta, Katelynn A. Kapalo, Andrew Best, and Stephen M. Fiore. 2016. Similarity, complementarity, and agency in HRI: Theoretical issues in shifting the perception of robots from tools to teammates. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60. SAGE Publications, Sage, CA: Los Angeles, CA, 1230–1234.
- [73] Auriel Washburn, Sachiko Matsumoto, and Laurel D. Riek. in press. Trust-aware control in proximate human-robot teaming. *Trust in Human-Robot Interaction* (in press).
- [74] Katherine S. Welfare, Matthew R. Hallowell, Julie A. Shah, and Laurel D. Riek. 2019. Consider the human work experience when integrating robotics in the workplace. In *Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 75–84.
- [75] Christopher D. Wickens and Stephen R. Dixon. 2007. The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science* 8, 3 (2007), 201–212.

Received June 2019; revised December 2019; accepted January 2020