Received: 20 May 2022 | Revised: 14 February 2023 | Accepted: 15 May 2023

DOI: 10.1002/hfm.20998

WILEY

Blaming yourself, your partner, or an unexpected event: Attribution biases and trust in a physical coordination task

Chi-Ping Hsiung | Gabriel A. León | David Stinson | Erin K. Chiou

Human Systems Engineering, Arizona State University, Mesa, Arizona, USA

Correspondence
Erin K. Chiou, Human Systems Engineering, Arizona State University, Santa Catalina Hall, Ste 150H, 7271 E Sonoran Arroyo Mall, Mesa, AZ 85212, USA.
Email: erin.chiou@asu.edu

Funding information
National Science Foundation, Grant/Award Number: OIA-1936997; Air Force Office of Scientific Research, Grant/Award Number: FA9550-18-1-0067

Abstract

As robots enabled by artificial intelligence become more agentic, people may come to develop trust schemas based on a robot's actions and attribute blame to the robot as they would with a human partner. Trust and blame have yet to be investigated during dynamic physical coordination tasks despite the potential ramifications for manufacturing and service industries that could benefit from effective human–robot physical coordination. In anticipation of future human–robot work configurations, we developed a joint physical coordination task as a preliminary test environment for understanding trust and blame in a work partner. Fifty-five participants were asked to jointly balance and transport a weighted box along a fixed path, and we used this test environment to evaluate the impact of a surprising event on trust in a work partner, and attribution of blame following a negative performance outcome. Results indicate that the group who experienced a surprising event compared to the group who did not trusted their partner more, but there was no difference in the attribution of blame to themselves, their partner, or to the surprising event. Conversely, the group who did not experience a surprising event tended to blame themselves for the negative outcome. These findings suggest that environmental uncertainty may prompt people's attribution of blame across multiple parties, including themselves. Moreover, people may build trust in work partners through the shared experience of surprising events. Future work would benefit from adopting our study design to investigate whether these findings are extendable to human–robot joint actors.

KEYWORDS
blame, human–robot, physical coordination, surprise, trust

1 | INTRODUCTION

Trust is a critical component in the workplace and due to the interdependence of tasks in many work environments, co-workers need to rely on one another to accomplish both individual and team goals (Mayer et al., 1995). Co-workers who trust one another will enjoy more effective team decision-making and proactive behavior in the workplace (Alge et al., 2003; Groysberg & Abrahams, 2006), which are important for organizational success in dynamic and competitive market environments. Although many definitions of trust have been used, an integrated review of the trust literature defines trust as the attitude that an agent will help with achieving a shared goal in situations characterized by risk and uncertainty (Lee & See, 2004).

Trust influences a person's willingness to rely on another agent during joint action. In trusting relationships, people will believe and come to expect that they can rely on their partners to help them meet their needs and goals (Campbell et al., 2010). These beliefs and

expectations can be informed by many factors related to a trustee's abilities and the trustee's reciprocation of care and concern for others (Al-Ani et al., 2012; McAllister, 1995). Within work teams, high trust implies confidence that a teammate will carry out what is expected of them in that team, without needing direct oversight (Sabherwal, 1999). Typically, higher trust frees up resources by reducing the need for excessive structural controls or monitoring within the work system, which can be costly in dynamic work environments. When team members trust one another, they can more easily negotiate actions to address dynamic factors in the task environment.

However, such trusting relationships may break when unexpected behaviors or events in the task environment increase risk or uncertainty. Such events may cause team members to question one another, ultimately leading to a breakdown of trust (Al-Ani et al., 2012). Moreover, organizational literature suggests that unexpected behaviors and surprises in the task environment can create opportunities for learning from positive or negative outcomes (Miner et al., 2001). In dynamic task environments, surprising events may introduce task conflict among teammates. Task conflict can occur when there is disagreement between team members, and when one team member perceives their interests are being opposed by another team member (Wall & Callister, 1995).

One way that team members cope with task conflict is through blame. Blame is used to make sense of difficult and complex situations (Kaniarasu & Steinfeld, 2014). People will attribute blame to help identify and control behavior so that those behaviors can align with social expectations (Malle et al., 2014). At the same time, avoiding blame can be self-serving, especially when people take credit for successes while denying any culpability for failures (Sedikides et al., 1998; Taylor & Doria, 1981). This self-serving bias is a cognitive reaction people use to protect themselves and this reaction can have different results depending on the proximity of the dyadic relationship (Sedikides et al., 1998). Because of the criticality of working relationships for joint task performance and the role that trust plays in those relationships, further understanding of how surprises in the environment can affect blame attribution and trust in work relationships should be explored.

1.1 | Research objective and practical motivation

The present study uses an experimental design to test the impact of a surprising event on trust and blame toward a work partner in a physical coordination task. The study was conducted in a laboratory environment to control for organizational factors that may conflate the impact of an unexpected event on trust and blame. This research adds the factor of surprise to Kim and Hinds (2006), using human–human dyads while assessing blame in a teaming context.

Although this study draws its theoretical foundation from the psychological sciences, the motivation for this work is also inspired by recent studies in human–robot work environments. One example is a field observation of an autonomous delivery robot in a hospital setting in which nurses were observed to blame co-workers for having done something to disturb the robot, when the robot produced inexplicable behaviors or errors (Hancock et al., 2011; Kaniarasu & Steinfeld, 2014; Mutlu & Forlizzi, 2008). Much of the ongoing work in robotics and human–robot interaction has been to better support the future of work in manufacturing and service industries, with attention to developing robots that can engage in joint activities with human counterparts. Yet, relatively little attention has been paid to the social implications of performing physical work jointly in dynamic task environments. This work context is a key area of development and an open area of research for integrating and deploying increasingly autonomous and collaborative robots.

2 | BACKGROUND

2.1 | Trust and attribution with work partners

Trusting another agent is a dynamic process that evolves over time and is influenced by people's perceptions of the situational context and the actors involved (Chiou & Lee, 2021). A precursor to trust development, then, is *attribution*. Attribution occurs when people create a causal explanation for behavior based on perceived information in their social environment (Heider, 1958; Moskowitz, 2005). An example of this can be seen in driving and road rage; people tend to blame others around them, especially for negative outcomes or experiences (Hancock et al., 2021). Additionally, people may attribute the cause of a negative outcome to situational or external factors, such as workplace-related uncertainty, or to dispositional or internal factors. When a negative outcome arises, a violation of trust may occur and cause the trustor to re-evaluate the current situation (Lewicki & Bunker, 1996). This supports the notion that attribution biases, including correspondence bias and self-serving bias, can direct peoples' assignment of blame. In research on revenge, people are found to make attributions immediately after a trust violation to help them decide how they should handle the situation (Bies et al., 1997). When it comes to working relationships, trust violations may result in *attribution bias* when certain factors influence people to attribute blame inaccurately (Manzey et al., 2012; Parasuraman & Riley, 1997).

People constantly make attributions for the cause of their own and others' behavior, yet people are also susceptible to perceptual errors. These perceptual errors can lead to biased interpretations of their social world (Funder, 1987) including self-serving bias. Self-serving bias is when a person tends to take undue credit for successes while denying any wrongdoings or failures (Taylor & Doria, 1981). Research has shown that a partner's actions are more salient targets of blame when compared to self-action or cues in the environment (Jones & Nisbett, 1987). Therefore, it is possible that when work partners coordinate, they are more likely to attribute negative outcomes to their partner rather than to themselves or to environmental factors (Walther & Bazarova, 2007).

In dynamic task environments, people calibrate their trust in other agents based on their attributions of blame and causality, but

these attributions may be biased. *Correspondence bias* is the tendency to attribute causality based on internal and dispositional characteristics of an agent, rather than to external factors (Gilbert & Malone, 1995; Jones, 1979). For example, when two agents work together, environmental instability may affect their partner's behavior, and people may overestimate the effects of dispositional factors and underestimate the effects of environmental instability on the outcome (Lassiter et al., 2002).

2.2 | Attribution biases and human–machine work relationships

Correspondence bias is identified as a potential explanation for inappropriate trust calibration (Wisse, 2010) and is shown to generalize to attributions of inanimate objects (Sharek et al., 2010). When it comes to interacting with machines, Muir (1987) argued that people are more likely to attribute unpredictability to a machine's properties over environmental instability, even when environmental instability is the main cause of the machine's behavior. However, it is possible that correspondence bias might simply be a problem of incomplete information. As an example of the many studies that demonstrate correspondence bias, Ross (1977) noted that it is difficult for people to precisely determine the strength of the relationship between cause and effect. More recent research suggests that correspondence bias can persist even when information about both the behavior and situation are known with equal clarity and are presented in the same format and modality (Moore et al., 2010).

When it comes to extending these findings to human–machine work relationships, previous research has found that people tend to blame technology for mistakes and errors while exhibiting reluctance to credit positive outcomes to technology (Friedman, 1995; Morgan, 1992; Sampson, 1986). This finding has been shown in studies of people interacting with a computer assistant (Moon & Nass, 1998), and with robots as instructors (You et al., 2011). In addition, self-serving bias is observed in tasks involving human–machine shared control. Vilaza et al. (2014) designed a computer game that required a machine agent and participant to jointly control the direction of a ball to avoid obstacles and collect a target item. Their findings indicate that participants would blame the machine agent when they lost a game, whereas they took credit when they won a game—evidence of self-serving bias.

2.3 | Attribution biases and trust in machines

When two or more agents work together to accomplish a task, these two attribution biases correspondence bias and self-serving bias may direct the way people interact with their partners—robotic or otherwise. For example, the development of interagent trust can be viewed as an attribution process. An individual may develop beliefs about another agent's trustworthiness based on whether the agent's behavior is judged to be caused by internal or external factors (Krosgaard et al., 2002). Ferrin and Dirks (2003) used attribution theory as a framework for understanding the multiple perceptual and behavioral routes through which reward structures—an omnipresent factor in work environments—influence trust development in a teammate. Moreover, does blame attribution impact subsequent trust in a work partner? Even within the research on human teammates, few studies investigate blame toward a teammate during joint physical tasks while simultaneously examining trust. The present study seeks to confront these questions by implementing a physical coordination task in a dynamic environment, that is, an environment that includes environmental instability.

2.4 | Unexpected events, surprise, and blame

Many system designers aim to mitigate the risk of being surprised in environmentally unstable conditions, by standardizing operator behavior and minimizing variability in the operational environment (Hollnagel, 2013). However, standardizing operator behavior and expecting operators to adhere to those standards at all costs can lead to brittle rather than flexible systems (Gomes et al., 2009; Rochlin et al., 1987). More pragmatically, eliminating surprises in environmentally unstable conditions may not be pragmatic for job environments with inherently high instability and risk of harm. Instead, operators can be trained to react quickly during narrow decision windows and re-establish interpredictability and common ground with their team members (Klein et al., 2005).

Surprises unexpectedly arise, yet are normal everyday occurrences in teams doing things (Stompff et al., 2016). Surprise is defined as a cognitive-emotional response to something unexpected, which results from a mismatch between one's mental expectations and perceptions of one's environment (Horstmann, 2006; Meyer et al., 1991; Schützwohl & Borgstedt, 2005). Appraisal theory emphasizes that *unexpectedness*, rather than novelty, unfamiliarity, or uncertainty, is the essential prerequisite to surprise (Roseman, 1996). Unlike *startle*, which occurs in response to a sudden, high-intensity stimulus, surprise can be evoked by an unexpected stimulus or by the unexpected absence of a stimulus (Rivera et al., 2014). During highly structured coordination tasks, surprise is often perceived as negative because it implies that the contingencies of successful coordination may be compromised (McDaniel et al., 2003).

To devise surprise-resilient systems, designers may need to consider the social and emotional significance of coordinated movement (Schmidt & Richardson, 2008). Physical synchrony and psychological synchrony are highly related (Marsh et al., 2009). Examples of surprises in sociotechnical workplaces can be found in human–machine coordination (Woods et al., 1997), aviation (Sarter & Woods, 1995), and operating rooms (Moll Van Charante et al., 1992). Indeed, successful coordination is more likely to occur when partners are aware of the other's intentions and trust that they will perform the expected movements in a proper sequence (Jacob & Jeannerod, 2005).

Yet, surprising events may cause a partner to doubt another's ability to adjust to the perturbations of an unstable system. With an unclear mental representation of their partner's intentions, a partner may be more likely to blame the other when performance goals are not met, rather than blame the overall system itself (Holden, 2009). Additionally, negative consequences are more likely to occur when there is a misalignment of trust and partners blame each other. To address this concern, the present study induced surprise during a joint physical coordination task to assess how partners distribute blame for poor performance and adjust their trust accordingly.

2.5 | Human–human interaction as a proxy for human–robot interaction

Joint physical coordination in human dyads and groups has been studied extensively. When observing specific motor patterns in others, people naturally build cognitive schemas for the movement that activate motor cortexes and mirror neuron systems (Rizzolatti et al., 2001). These action representations within the motor cortexes help to simulate potential behavioral responses in partners (Jeannerod, 2001), anticipate intentions and goals (Cuijpers et al., 2006), and enact movements that meet the spatial and temporal demands of the coordination task (Sebanz & Knoblich, 2009). Interestingly, similar mechanisms for deciphering motor patterns in human partners have been implicated in evaluating movement in nonhuman agents. Under functional magnetic resonance imaging screening, observing anthropomorphic movement and robotic movement generated activity in very similar, yet distinct, brain regions (Kuz et al., 2015). An overlap in neural stimulation across various agents supports the need to establish research that examines human interactions in situations that might extend to human–robot coordination.

From the research that is available, example domains in which robots are being used in close coordination with people include: space exploration, search and rescue, surgery, assisting older adults and people with disabilities in daily living activities, and manufacturing (Heerink et al., 2010; Hinds et al., 2004; Onnasch & Hildebrandt, 2022; Parasuraman et al., 2009). The combined efforts of human–human and human–robot research will need to be strengthened so that the ongoing development of human–robot physical coordination tasks can be successful in the future (Hancock et al., 2011).

As robots become more capable, and more autonomous, they may come to take on more responsibility within certain tasks parameters when working with a human partner (Kaniarasu & Steinfeld, 2014). However, with this increasing robot capability comes the need to better understand the conditions of trust between physically coordinating work partners. Trust in a work partner—robotic or otherwise—is an indicator of how well those agents would work together in the future (Freedy et al., 2007). But when surprises occur, followed by a negative performance outcome, people may find themselves blaming the robot to make sense of the situation and cause a breakdown in the trust relationship. Therefore, one pressing question is whether people will exhibit specific attribution biases that will affect their perceived trustworthiness of a work partner, especially during joint physical coordination tasks in which instances of environmental instability are experienced.

From a behavioral perspective, human dyads may be used to further understand these social changes that people go through to learn how the robot should operate. Organizations rely on teams to get work done, and robots are entering the team atmosphere (Jung et al., 2017). A better understanding of human–human coordination tasks could inform the design of robots that can coordinate more naturally with people (De Santis et al., 2008). Similarly, social interactions between people can be studied to inform the design of more human-friendly robots (Cassell & Bickmore, 2000). As robots begin to possess more advanced social capabilities, they too will have the ability to respond to changes, communicate intentions, and request action during teaming tasks (Chiou & Lee, 2021; Jung et al., 2017; Wortham et al., 2016).

2.6 | Background summary and hypotheses

To summarize, the present study examines how trust develops after experiencing an unexpected surprise-inducing event during a physical coordination task. We also explore the potential relationship between attribution biases and trust. From previous literature, we know that under certain conditions, people are known to blame a negative outcome on their partner and are less likely to blame themselves (i.e., self-serving bias). Furthermore, we know that people tend to consider that their partner's behavior is due to their partner's internal characteristics rather than external factors that the partner might be facing (i.e., correspondence bias). Drawing from appraisal theory, we can surmise that when coordinating with a partner on a task that (1) encounters an unexpected (i.e., surprising) event and (2) results in a negative outcome, people will tend to blame the outcome on their partner compared to the same coordination task without an unexpected event, because people would attribute their partner's internal characteristics as the primary cause of the negative result. The specific hypotheses for this study are as follows:

H1. Effect of attribution biases and environmental instability on attribution of blame:

a. Participants in a no-surprise (baseline) condition will tend to blame their partner rather than themselves for a negative outcome.
b. Participants in a surprise condition will tend to blame their partner for a negative outcome, above-and-beyond any blame directed toward themselves or the surprise event.

H2. Effect of environmental instability on trust:

a. Participants in the no-surprise (baseline) condition will have similar pre- and posttask trust measures, and similar measures

of predictability, dependability, reasonableness, or competence (i.e., known factors that inform trust).

b. Participants in the surprise condition will show decreases in pre- to post-task trust, and decrease in measures of predictability, dependability, reasonableness, and competence (i.e., known factors that inform trust).

H3. Effect of partner blame on posttask trust: Partner blame will predict decreases in posttask trust from the no-surprise to the surprise conditions.

3 | MATERIALS AND METHODS

Fifty-five undergraduate students (17 women and 38 men) from a large university in the southwestern United States participated in this study, which was approved by the university's Institutional Review Board that evaluates the ethical conduct of planned research activities. Participants were recruited through an online course credit management system, paper flyers, and in-person solicitation for volunteers. All participants reported they had not met their partner before the study, were able to carry 10 pounds with their dominant arm, and were comfortable communicating in English. All participants were required to be at least 18 years old. Participants recruited from the online course credit management system received credits toward an assignment that is a small part of their grade in a course.

3.1 | Procedure

Upon arrival, two researchers provided each participant with a brief overview of the study and asked them to read and sign an informed consent form, complete a demographic questionnaire, and an initial trust questionnaire (Merritt & Ilgen, 2008; Muir, 1987). Participants were randomly paired and assigned to one of two conditions (surprise and no-surprise) that involved completing a simple joint physical coordination task (i.e., lifting a weighted box). This task was designed to function as a simplified version of a real work system in which the essential elements are retained and the complexities eliminated to make experimental control possible (Brehmer & Dörner, 1993). The "no-surprise" baseline condition simply involved completing the task, and the "surprise" test condition involved completing the task with an interjecting warning tone that periodically interrupted the task. The same instructions for the task were provided to all dyads, along with guidelines for how to respond to the warning tone *should they hear one*.

After researchers confirmed that participants were comfortable with the task instructions, participants completed one practice trial before proceeding to four experimental trials. The task began with the box on the ground within a square marked with blue-colored tape (Figure 1a). Another square was marked on the table to indicate where participants were to move the plastic box (Figure 1b). Two other squares were marked on the ground with the numbers "1" or "2," indicating where the participants were instructed to stand during

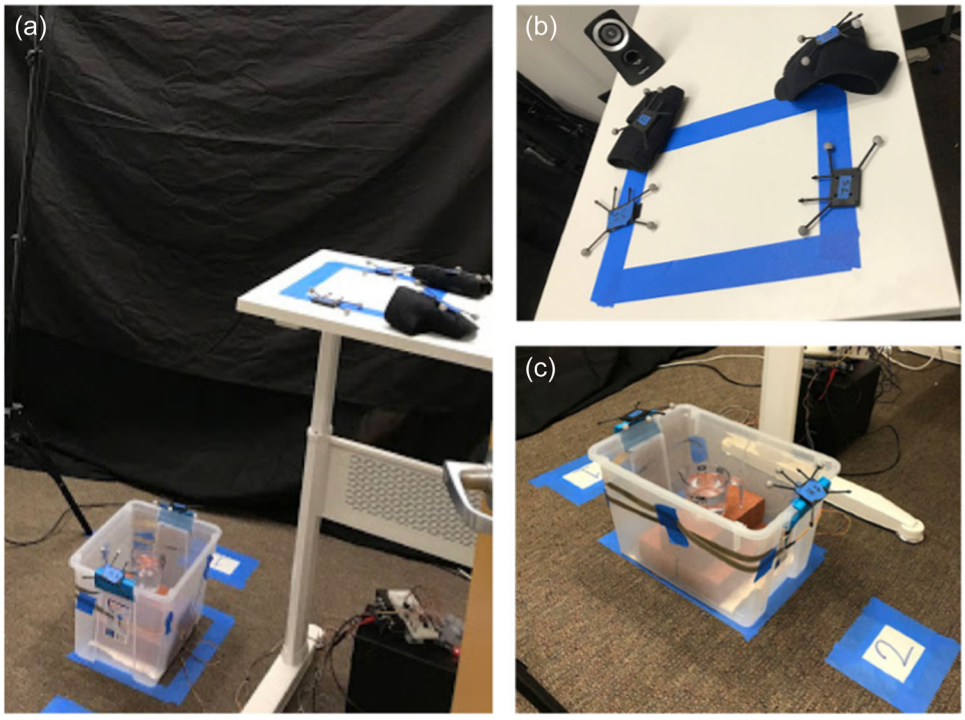

FIGURE 1 Experimental task room setup. The image shows three different angles of the experimental room setup. Clockwise from left: (a) A translucent box on the floor right next to a table; (b) a close-up of the table-top, with a speaker and motion sensors placed just beyond a blue-taped rectangle; (c) a close-up of the box on the floor with motion sensors attached to the sides of the box.

each trial (Figure 1c). The purpose of these markings was to help clarify the task instructions, and also aid in motion-sensor capture (sensors seen throughout Figure 1). Data collected from these sensors were not analyzed in this study but were used to inform a different study on robot motion planning.

After completing all trials, and regardless of performance, participants were informed that based on their performance, they did not win the prize (see next section for more detail). Lastly, participants were asked to complete a trust and blame questionnaire. The overall procedure took less than 1 h to complete.

3.2 | Coordination task instructions

Participants were asked to lift a translucent box with their partner from the designated area on the ground to a table (90 cm in height), then return the box back to the ground, which completed one trial. Participants completed a total of four trials. The box contained three bricks weighing 5.75 lbs in total and on top of those bricks was a cup filled with 200 mL of water. During the transportation task, participants were told that they would be graded on their ability to prevent water spillage and ability to complete the task in the shortest time. These conditions simulate demands on speed and accuracy for physical work within a compensatory reward structure that applies to many other domains of joint physical coordination (Braam et al., 1996).

Participants were informed that the team with the best performance would receive a $30 prize in the form of a Starbucks gift card. While the coordination task was reward incentivized, regardless of their true performance, all groups were fixed to receive a negative performance evaluation (i.e., told that they did not overcome the highest scoring group) to prompt a reason for holding some party blameworthy for poor performance (Ferrin & Dirks, 2003).

3.3 | Surprise warning tone

To simulate an unexpected event, during the surprise condition, a 250 Hz tone was played via Logitech Z313 speakers for 4 s, in the second and fourth trials. This means that the surprise group would lift the box a total of four times with the same partner each time, but they would only hear the tone twice, once in the second trial and once in the fourth trial. The tone was coded to trigger 0.8 s after the dyad lifted the box from the ground, for both trials. Participants were notified before starting the task that a warning tone *might* occur but *not when* or if it would occur. They were instructed to hold the box still while the tone sounded, and then to resume moving the box to complete the trial once the tone ceased.

The purpose of holding still was to simulate a situation in which (1) their task performance is affected by something in the environment and (2) they must physically coordinate to minimize negative impacts on the outcome, that is, working together while experiencing an unexpected event. If the dyads are able to keep still during the sounding of the tone, they are indicating their awareness of the current situation and expectations. This also simulates a situation in which dyads are working together, and if both partners are not aligned to their task environment and potential perturbations, then the risk of engaging in that task with another person is much higher to human safety and performance in the workplace. Therefore, although participants knew about the possible existence of the warning tone and were given specific task-impacting instructions for what to do should they hear one, they were not aware of if, when, or how often the tone would sound. This part of the task was designed to create a *surprise*-inducing event and not a *startle*-inducing event—aligning with the framework outlined by Kochan et al. (2005), which defines an unexpected event as:

> An event incongruent with expectations as determined by base rate probabilities (average probability of event occurring) and the contextual information available; may be normal, abnormal, or emergency in nature; it may also be frequent, infrequent, or novel. (p. 339)

Therefore, surprise instead of startle was implemented not only to minimize potential injury to participants but also because understanding potential human responses to surprise will be important for designing robust robot partners that are being built to engage in human–robot physical coordination tasks in dynamic environments.

3.4 | Measures

3.4.1 | Partner trust

Participants' perceptions of trust in their partner were obtained on a 5-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree), using an adapted instrument from Merritt and Ilgen (2008). In our adaptation, we used five items total and changed the original referent used in their study to the word "partner." One item assesses the perceived overall trustworthiness of a partner, and the other four items each relate to the trust-related factors identified in Muir (1987), which include predictability, dependability, responsibility, and competence. Although there are many "correct" ways to measure trust (Kohn et al., 2021), we chose the instrument used by Merritt and Ilgen (2008) as a starting point due to the instrument's conciseness. Rather than selecting an instrument that would help us tease apart specific dimensions of trust, as might be the case for choosing a longer instrument like Chancey et al. (2017), we primarily cared about general trust perceptions following our experimental manipulations. We were also confident that the four items would sufficiently capture the various signals of trust that might be present in our task environment, given the repeated observations in other studies that have shown positive relationships between these items and trust, particularly in studies involving perceptions of people and machines (Lee & See, 2004).

This questionnaire was administered twice, once pretask and again posttask. The pretask questionnaire was administered after the

training slides to measure participants' initial trust in their partners. The second was administered after the conclusion of the last trial. Scale reliability was acceptable for the no-surprise baseline (pretask, $\alpha = .89$; posttask, $\alpha = .96$) and surprise condition (pretask, $\alpha = .81$; posttask, $\alpha = .79$). See Table 4 for a summary of descriptive statistics.

3.4.2 | Attribution of blame

To assess participants' attribution of blame, a questionnaire administered on Qualtrics (an online survey tool) was adapted from Kim and Hind's (2006) study on the attribution of blame and credit toward people and robots. Participants were asked two questions about the level of blame and responsibility they would ascribe to themselves, their partner, and the warning tone, for the negative outcome. These questions were answered on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Final scores were derived from averaging the two questions for each of the three targets of blame. The final scores for self-blame, partner blame, and warning-tone blame were used to predict intergroup variability in our outcome measures. Scale reliability was acceptable for baseline (self-blame, $\alpha = .92$; partner blame, $\alpha = .89$, warning-tone blame, N/A) and surprise conditions (self-blame, $\alpha = .97$; partner blame, $\alpha = .90$; warning-tone blame, $\alpha = .97$). See Table 1 for a summary of descriptive statistics.

3.4.3 | Surprise

Surprise, not startle, can be measured by self-report or behavioral methods (Loewenstein, 2019). We used a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree) as a self-report measure of surprise (Reisenzein et al., 2006). Participants who reported four or higher on the self-report surprise scale were considered surprised. Researchers recorded each participant's response after the warning tone was played.

A facial expression checklist was also used as a behavioral measure of surprise. Participants' facial expressions were documented by researchers while the warning tone was playing. Any participant who showed at least one facial expression was considered surprised (Ekman & Rosenberg, 1997). The self-report and facial expression checklist were used to validate the manipulation of surprise in the test group.

4 | RESULTS

Sixty participants were recruited in total, with 30 unique participants randomly assigned to each of the two test conditions. However, only 25 participants in the no-surprise (baseline) condition were included in the final analyses. Two pairs of participants did not receive the trust questionnaire due to a procedural error, and one participant opted not to answer the trust questionnaire entirely. All 30 participants from the surprise condition were included. Therefore, the final sample size across both groups sums to 55.

4.1 | Test diagnostics for attribution of blame

Self-blame, partner blame, and warning-tone blame scores were tested for normality. In the no-surprise (baseline) condition, scores for self-blame ($W = 0.87$, $p < .01$), partner blame ($W = 0.83$, $p < .01$), and the difference values between self-blame and partner blame ($W = 0.813$, $p < .01$) were significantly nonnormal. In the surprise condition, scores for self-blame ($W = 0.9$, $p < .01$), partner blame ($W = 0.882$, $p < .01$), and warning-tone blame ($W = 0.856$, $p < .01$) were also significantly nonnormal. Warning-tone blame scores were excluded in the no-surprise condition because they did not apply.

Variances were dissimilar between self-blame and partner blame in the no-surprise condition ($F(1,48) = 11.255$, $p < .01$), and across self-blame, partner-blame, and warning-tone blame in the

TABLE 1 Descriptive statistics of attribution of blame questionnaire.

Scales	Baseline ($n = 25$)			Surprise ($n = 30$)		
	Median	*M*	SD	Median	*M*	SD
Attribution of blame to self						
I was responsible for the unsuccessful result	3.00	2.88	1.81	3.00	2.93	1.76
I was responsible for the unsuccessful result	2.00	2.76	2.03	3.00	2.90	1.81
Attribution of blame to the partner						
My partner was to blame for the unsuccessful result	2.00	2.32	1.46	2.50	2.57	1.59
My partner was responsible for the unsuccessful result	1.00	2.16	1.41	2.00	2.37	1.47
Attribution of blame to the warning tone						
The warning tone was to blame for the unsuccessful result	N/A	N/A	N/A	3.00	3.17	2.23
The warning tone was responsible for the unsuccessful result	N/A	N/A	N/A	3.00	3.10	2.14

Note: Scales for all range from 1 (strongly disagree) to 7 (strongly agree).
Abbreviation: NA, not available.

TABLE 2 Spearman's correlations in baseline condition ($n = 25$).

	1	2	3	4	5	6	7	8	9	10	11	12
1. Pretask predictability	–											
2. Pretask dependability	0.572*	–										
3. Pretask reasonability	0.47	0.81*	–									
4. Pretask competence	0.288	0.653*	0.708*	–								
5. Pretask overall Trust	0.646*	0.815*	0.758*	0.647*	–							
6. Posttask predictability	−0.21	−0.08	−0.029	0.023	0.096	–						
7. Posttask dependability	−0.102	0.127	0.149	0.182	0.337	0.708*	–					
8. Post-Task Reasonability	−0.069	0.158	0.175	0.158	0.37	0.671*	0.902*	–				
9. Posttask competence	−0.178	0.132	0.23	0.224	0.304	0.765*	0.952*	0.844*	–			
10. Posttask overall trust	0.037	0.281	0.195	0.188	0.538*	0.588*	0.733*	0.846*	0.692*	–		
11. Self-blame	0.015	0.146	0.205	0.09	0.067	0.394	0.076	0.051	0.18	0.233	–	
12. Partner blame	−0.034	−0.003	0.017	−0.044	−0.174	−0.021	−0.301	−0.289	−0.219	−0.178	0.532*	–

*$p < .01$.

TABLE 3 Spearman's correlations in surprise condition ($n = 30$).

	1	2	3	4	5	6	7	8	9	10	11	12
1. Pretask predictability	–											
2. Pretask dependability	0.078	–										
3. Pretask reasonability	0.365	0.716*	–									
4. Pretask competence	0.202	0.595*	0.642*	–								
5. Pretask overall trust	0.393	0.333	0.55*	0.482*	–							
6. Posttask predictability	0.384	0.045	0.258	0.296	0.275	–						
7. Posttask dependability	−0.065	0.413	0.405	0.564*	0.204	0.298	–					
8. Posttask reasonability	0.214	0.043	0.357	0.318	0.311	0.393	0.666*	–				
9. Posttask competence	0.21	0.224	0.489*	0.382	0.44	0.59*	0.654*	0.751*	–			
10. Posttask overall trust	0.365	0.159	0.491*	0.486*	0.664*	0.133	0.412	0.647*	0.526*	–		
11. Self-blame	−0.057	0.16	−0.143	0.039	−0.003	−0.087	0.067	−0.105	−0.267	−0.237	–	
12. Partner blame	−0.097	−0.15	−0.179	−0.19	−0.344	−0.388	−0.199	−0.081	−0.301	−0.158	0.427	–
13. Warning-tone blame	−0.088	0.167	−0.028	0.127	−0.279	−0.47*	0.006	−0.066	−0.288	−0.041	0.419	0.849*

*$p < .01$.

surprise condition ($F(2,87) = 3.12$, $p = .05$). Nonparametric repeated-measures tests were used to test mean differences between the targets of blame. Spearman's correlations are reported in Tables 2 and 3.

4.2 | Impact of surprise on attribution of blame

Wilcoxon's signed-rank test was used to compare self-blame and partner blame in the no-surprise condition. Partner-blame scores ($M = 1.76$, median [Mdn] = 1.5) were significantly different and lower than self-blame scores ($M = 2.82$, Mdn = 2.5; $W = 112.0$, $p < .01$, $r = -.42$). Therefore, hypothesis H1a was not supported.

A Friedman analysis of variance (ANOVA) was conducted to compare self-blame ($M = 2.92$, Mdn = 2.75), partner blame ($M = 2.77$, Mdn = 3.00), and warning-tone blame scores ($M = 3.13$, Mdn = 2.75) in the surprise condition. There was no significant difference among the three blame sciores ($\chi^2(2) = 2.97$, $p > .05$). Therefore, H1b was not supported.

There was no significant difference in self-blame across no-surprise ($M = 2.82$, Mdn = 2.50) and surprise conditions ($M = 2.92$, Mdn = 2.75; $W = 360.5$, $p < .81$, $r = -.03$). However, partner blame was significantly higher in the surprise condition ($M = 2.77$, Mdn = 3.00) when compared to the no-surprise condition ($M = 1.76$, Mdn = 1.50; $W = 223.0$, $p < .01$, $r = -.35$). At this point, we could tentatively say that H3 was partially supported in terms of detectable differences in blame, but not in trust. Table 1 illustrates descriptive statistics for blame attributions across test conditions, and Figure 2 illustrates these comparisons.

4.3 | Test diagnostics for trust

In the no-surprise condition, overall pretask trust ($W = 0.80$, $p < .01$) and posttask trust ($W = 0.73$, $p < .01$) were significantly nonnormal, and in the surprise condition, overall pretask trust ($W = 0.78$, $p < .01$) and posttask trust ($W = 0.68$, $p < .01$) were also significantly nonnormal. Moreover, in the baseline condition, difference scores between overall pre- and posttask trust were significantly nonnormal ($W = 0.85$, $p < .01$). Difference scores were also significantly nonnormal in the surprise condition ($W = 0.74$, $p < .01$). Variances were not significantly different between overall pre- and posttask trust in the baseline ($F(1,48) = 0.34$, $p = .56$) and surprise conditions ($F(1,58) = 3.63$, $p = .06$).

In the no-surprise condition, posttask scores for perceived partner predictability ($W = 0.76$, $p < .01$), dependability ($W = 0.76$, $p < .01$), reasonability ($W = 0.79$, $p < .01$), and competence ($W = 0.80$, $p < .01$) were all significantly nonnormal. Likewise, in the surprise condition, ratings for partner predictability ($W = 0.69$, $p < .01$), dependability ($W = 0.65$, $p < .01$), reasonability ($W = 0.55$, $p < .01$), and competence ($W = 0.62$, $p < .01$) were all significantly nonnormal. Nonparametric tests were used to assess the impact of surprise on each subcomponent of trust. Correlations are reported in Tables 2 and 3.

4.4 | Impact of surprise on pre- and posttask trust in partner

Overall pretask trust ($M = 3.96$, Mdn = 4.00) and overall posttask trust scores ($M = 4.12$, Mdn = 5.00) were not significantly different in the no-surprise condition ($W = 27.0$, $p = .34$, $r = -.13$). Therefore, H2a was supported. However, pretask trust ($M = 4.03$, Mdn = 4.0) and posttask trust ($M = 4.47$, Mdn = 5.0) were significantly different in the surprise condition ($W = 4.0$, $p < .01$, $r = -.34$), with posttask trust being significantly higher than pre-task trust. Therefore, H2b was partially supported in that there is a difference, but not in the direction we expected. For context, there was no significant difference in overall post-task trust across no-surprise ($M = 4.12$, Mdn = 5.00) and surprise conditions ($M = 4.47$, Mdn = 5.00; $W = 334.5$, $p < .44$, $r = -.10$). Figure 3 illustrates these findings through visual comparison.

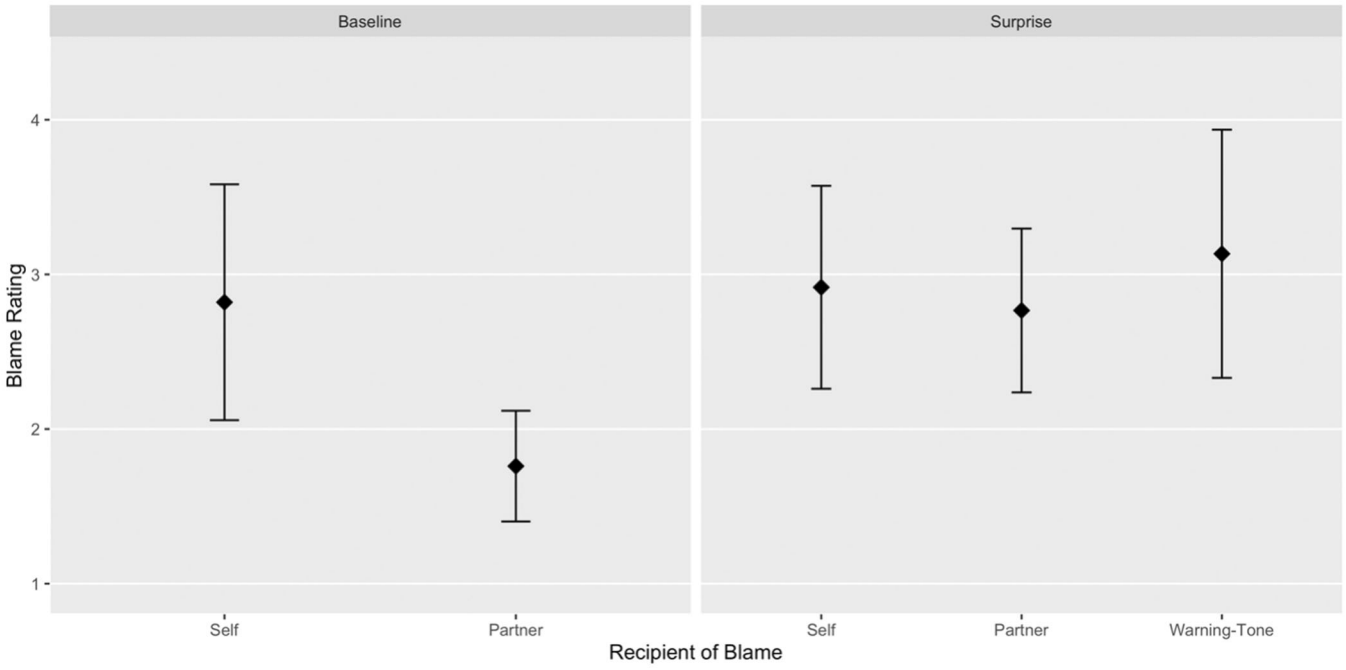

FIGURE 2 Average blame in no-surprise (baseline) and surprise conditions. In the no-surprise (baseline) condition with no warning tone ($n = 25$), participants blamed themselves more than their partners for the negative outcome. In the surprise condition ($n = 30$), blame for the negative outcome was evenly distributed among self, partner, and warning tone. Average ratings (from a scale of 1–7) are shown with 95% confidence intervals in each condition.

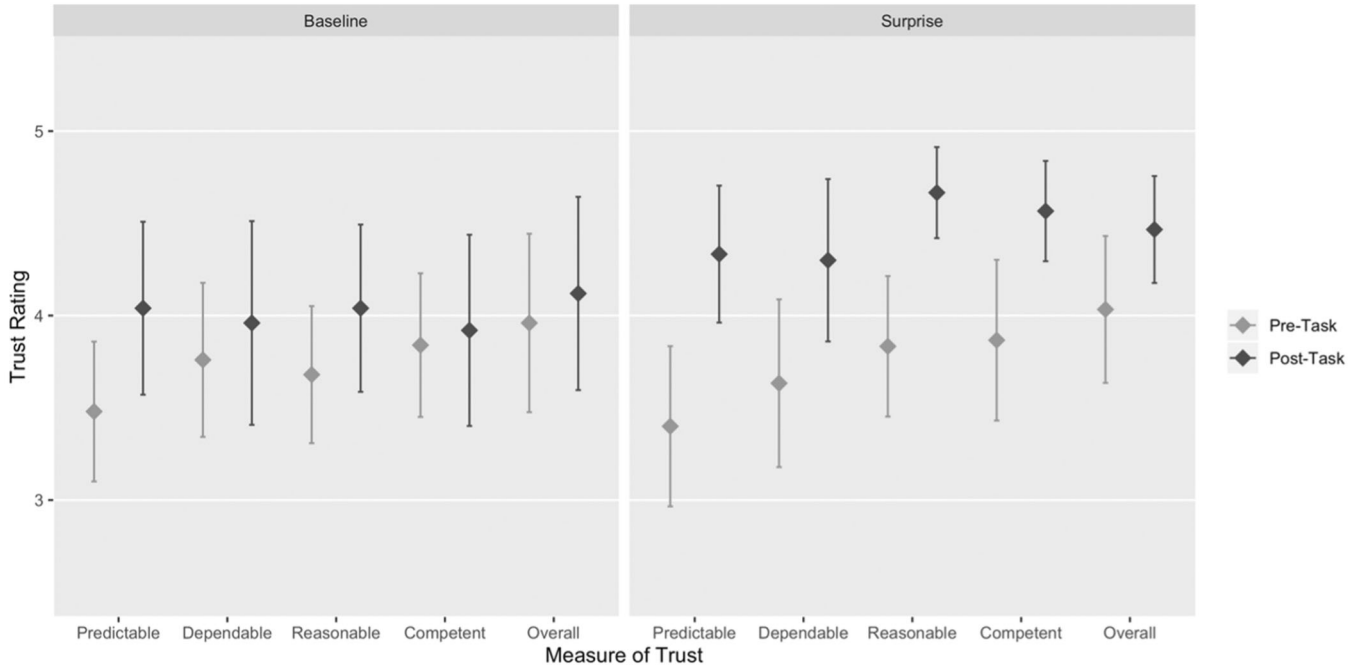

FIGURE 3 Average pre- and posttask trust ratings. Average trust responses that were rated on a scale of 1–5. Error bars shown for each trust questionnaire item are 95% confidence intervals.

TABLE 4 Muir's trust questionnaire ratings across baseline and surprise conditions.

	Baseline (n = 25)						Surprise (n = 30)					
	Pretask			Posttask			Pretask			Posttask		
	M	Median	SD	*M*	Median	SD	*M*	Median	SD	*M*	Median	SD
Overall trust	3.96	4.00	1.18	4.12	5.00	1.27	4.03	4.00	1.07	4.47	5.00	0.78
Predictability	3.48	3.00	0.92	4.04	4.00	1.14	3.40	3.00	1.16	4.33	5.00	1.00
Dependability	3.76	4.00	1.01	3.96	4.00	1.34	3.63	4.00	1.22	4.30	5.00	1.18
Reasonability	3.68	4.00	0.90	4.04	4.00	1.10	3.83	4.00	1.02	4.67	5.00	0.67
Competence	3.84	4.00	0.94	3.92	4.00	1.26	3.87	4.00	1.17	4.57	5.00	0.73

Note: Scales range from 1 (strongly disagree) to 5 (strongly agree).

4.4.1 | Impact of surprise on perceived partner predictability, dependability, reasonability, and competence

In the no-surprise condition, there were no significant differences between pre- and posttask perceived partner predictability (W = 40.5, p = .09, r = −.24), dependability (W = 70.0, p = .30, r = −.15), reasonability (W = 43.0, p = .10, r = −.23), or competence (W = 80.5, p = .55, r = −.08). Therefore, H2a was supported. However, in the surprise condition, partners were rated as significantly more predictable (W = 9.0, p < .01, r = −.45), dependable (W = 26.0, p < .01, r = −.37), reasonable (W = 6, p < .01, r = −.47), and competent (W = 24.0, p < .01, r = −.41) in the posttask assessment when compared to their pretask ratings. Therefore, H2b was again partially supported in that there was a difference but not in the direction we expected.

For context, perceived posttask ratings of predictability (W = 308.5, p = .22, r = −.16) and dependability (W = 315.5, p = .27, r = −.15) were not significantly higher in the surprise condition compared to the no-surprise condition. However, posttask ratings of partners' reasonability (W = 238.0, p < .01, r = −.36) and competence (W = 264.0, p < .05, r = −.28) were significantly higher in the surprise condition. Table 4 shows descriptive statistics for overall trust, predictability, dependability, reasonability, and competence.

4.5 | Manipulation checks

Descriptive data from the self-report and facial expression checklist measures of surprise suggest that participants in the surprise condition were often surprised by the first warning tone, but not the second. During the first warning tone, 79.16% and 62.5% of

TABLE 5 Facial expression checklist for surprise and self-reported surprise.

	Facial expression checklist		Self-report surprise scale	
	No	Yes	<4	≥4
First warning tone	5	19	9	15
Second warning tone	13	11	17	7

Note: Six characteristics were used to analyze the facial expression, which was the movement of their eyebrows, eyes, and jaw, if there are any sudden movements, if there are any sudden noises, if participants look surprised, and if they gazed in the direction of their partner. Any participant who showed at least one change in facial expression following the warning tone was marked as surprised. Self-report surprise scale ranged from 1 (not at all surprised) to 7 (as surprised as one can be). Any participant who reported 4 or higher on the self-report surprise scale was considered surprised.

participants qualified as surprised in the facial expression and self-report measures, respectively. Only 45.83% and 29.16% were surprised by the same measures during the second warning tone. Table 5 summarizes the facial expression checklist and self-reports of all participants. Six participants were not included in these tables due to data collection errors (missing or incorrectly measured).

Regression with bootstrapped coefficients and confidence intervals ($R = 20000$) was used to accommodate violations of assumptions for posttask trust. The surprise condition was a significant predictor of posttask trust ($b = 0.60$, SE = 0.26, 95% confidence interval [CI] = 0.11, 1.15), but partner blame did not predict posttask trust ($b = -0.14$, SE = 0.09, 95% CI = −0.32, 0.03). After adding an interaction term for the surprise condition and partner blame, none of the effects on posttask trust were significant for the surprise condition ($b = 0.39$, SE = 0.58, 95% CI = −0.58, 1.76), partner blame ($b = -0.22$, SE = 0.27, 95% CI = −0.78, 0.31), or the interaction term ($b = 0.10$, SE = 0.29, 95% CI = −0.45, 0.68). Therefore, overall H3 is not supported. Although the surprise condition predicts posttask trust, descriptively speaking it did not result in lower trust, and the effect is unlikely because of increased partner blame in our test environment.

5 | DISCUSSION

This study examined the impact of an unexpected, surprise-inducing event on trust and blame during a physical coordination task. To motivate blame, dyads were told that they had not performed the task well enough to win an anticipated prize. We predicted that surprise would lower trust and increase blame toward partners. This accords with literature on attribution biases that show people tend to blame negative outcomes on the internal characteristics of others (Gilbert & Malone, 1995; Jones, 1979) while deflecting blame away from themselves (Davis & Davis, 1972; Mezulis et al., 2004; Miller & Ross, 1975). The current human–robot interaction literature supports attributional bias and suggests people may blame robots for failures and take credit for positive actions (Groom et al., 2010). The goal of this study was to design a task environment that could generate insights from a human–human study to inform future studies of human–robot joint action. It is important for human–robot interaction research to continually unravel what human–robot relationships may look like in the future and what characteristics of joint work are unique to human–robot work relationships versus human–human work relationships.

The hypotheses regarding both biases were not supported (H1a and H1b). Evidence in support of the correspondence and self-serving biases was limited. Partner blame was significantly lower than self-blame in the baseline condition, but partner blame and self-blame were not different in the surprise condition. This indicates that participants were not likely to blame partners more than they would blame themselves in either condition. This might be because self-serving bias in causal attributions can be attenuated when people perform in groups (Zaccaro et al., 1987), and during face-to-face interaction (Glaeser et al., 2000), which was the case in our study. When people are collocated working together, there appears to be less blame on partners because both parties can sense the current situation (Walther & Bazarova, 2007).

One other reason for low partner blame could be that there was no punishment for failure. Nothing was taken away, but rather a potential prize was not awarded. In work environments suffering from blame culture (Timms, 2022), and when workers risk losing something they value highly (i.e., face, reputation, job), the attribution of blame to others may be much higher than when workers perceive that they have nothing to lose. Nonetheless, it seems that being surprised did cause partner blame to increase, even if the increase in partner blame did not surpass self-blame. Namely, the tendency for participants to blame themselves more than their partner in the baseline condition was nullified in the surprise condition—and blame was more evenly distributed to self, partner, and the warning tone.

All trust metrics (i.e., overall trust, predictability, dependability, reasonability, competence) remained consistent before and after the task in the baseline condition (H2a), whereas they increased in the surprise condition (H2c). These findings may be an artifact of the reward structure that we implemented—good performance was rewarded, and poor performance was not penalized. This structure was intentionally selected to avoid unnecessarily exposing our study volunteers to an unpleasant situation (penalizing them for poor performance), which we felt would not outweigh the benefits of the research. However, this does not detract from the fact that many work environments today still follow reward structures that penalize for poor performance, and in these situations, high partner trust in dynamic task environments may not persist.

Contrary to initial predictions, encountering a surprise-inducing event bolstered trust in a partner during a physical joint task, and marginally increased blame toward all actors (self, partner, environment) in the system. The trust increase may have been the result of social solidarity that formed between the human partners during the more challenging task, compared to the no-surprise condition. Social solidarity may have been compounded by in-group identification as

well; all participants were students from the same large university despite not knowing one another coming into the task. Perhaps because participants were able to witness how their partner responded to the warning tone, their confidence in their capabilities for future cooperation increased, despite not winning the prize (Lewicki et al., 2006). Bootstrapped regression analyses confirmed that people in the surprise condition were more likely to trust their partner after the task. Partner-blame and posttask trust were not significantly related.

While the warning tone did cause participants to blame their partners more on average, they were just as likely to blame themselves for their performance. Considering that the partner did not commit a violation of trust, it seems that the surprising event itself captured some of the blame that the participants would have directed toward themselves, as they did in the baseline condition. This supports the initial premise that surprise events may cause people to re-evaluate a priori attributions of blame (Kim et al., 2006). Irrespective of blame, participants seemed more willing to trust their partners and see them as more predictable, dependable, competent, and reasonable.

These observations offer a novel analysis of trust, blame, and surprise in a joint physical coordination task, in the hope that these methods will be tested in human–robot dyads. Prior research has demonstrated that social rules guiding human–human interaction may also apply to human–computer interaction, with people responding to machines as independent entities rather than as a manifestation of their human creators (Sundar & Nass, 2000). Additionally, and not tested in this research, it may be possible that as people work with robots more, they come to sense the robot as a teammate and do not blame or take credit from the robot. Similarly, there is evidence that people tend to treat robots as social actors and robots are not always perceived as mere tools (Friedman et al., 2003; Lee et al., 2005; Young et al., 2009). Altogether, these studies indicate that social psychological theory and simulated team task environments can enlighten our understanding of how people interact with robots (Cassell & Bickmore, 2000). Likewise, this study provides grist for further inquiry into human–robot teaming.

5.1 | Limitations

We acknowledge several limitations of the present study. First, the sample consists primarily of male-skewed, college students between 18 and 23 years old. We suggest that future researchers collect data from a larger and more diverse sample to generalize findings to operators that may be working alongside robots in physical coordination tasks in various settings. The second limitation involves the motivation for performance. We attempted to standardize a motive for effortful participation by offering a $30 value prize to participants with the best performance. However, this reward structure employs positive reinforcement only, absent the incentive to avoid scrutiny or punishment for failing to perform. Other types of performance incentives (e.g., punishment aversion, preservation of reputation, risks to safety, etc.) are likely to be relevant for understanding the effects of blame and trust in similar human–robot work structures.

The third limitation stems from the joint task itself. The act of moving a box in a controlled laboratory environment was designed for experimental control and generalizability, but will likely fail to encompass the many unique instances of human–robot physical coordination in unstructured environments (Ajoudani et al., 2018; Glasauer et al., 2010). Additional research can garner a more refined view of the dynamics of trust in human–robot teams by employing more domain-specific job tasks, interaction structures, and physical environments.

Similarly, a fourth limitation is the use of the warning tone as an analog for an unexpected event that elicits surprise. Despite precedence in prior research and overlap with real-world applications, for example, stall warnings in aviation systems (Landman et al., 2017), a warning tone does not generalize to all surprise-inducing events. The facial expression checklist and self-report checks of the surprise manipulation indicate that a majority of participants were surprised by the first tone, but fewer were surprised by the second tone—likely resulting from a habituation paradigm (McDaniel et al., 2003).

Though the facial checklist observes standardized behavior, it might misjudge people who do not express surprise through articulated facial movements (Reisenzein et al., 2006). Equally, self-report measures limit responses to post hoc reflection that may miss in-the-moment behaviors, feelings, and thoughts (Takayama, 2009). A participant might reflect and believe they were not surprised, but their behaviors or reaction time in the task could still indicate that the warning tone affected their response. Furthermore, it is possible that people define surprise differently than technical definitions of surprise and startle within the academic vernacular. Perhaps, some participants judged themselves to be *not* surprised because they remembered that they were notified that a tone might occur and were trained on what to do if/when they heard the tone. In future studies, electroencephalograms, skin conductance, and other measures of biological markers of surprise may enhance efforts to validate manipulations of surprise.

Finally, the present study uses an all-human dyad to infer an understanding of how people may respond to a robot partner. Though current research shows that robots with humanoid appearance and motor patterns may engage the same neural pathways as human–human interactions (Glasauer et al., 2010), many robots lack humanoid attributes and thus cannot be expected to stimulate the same social responses as a human partner. Future studies can continue this line of inquiry by directly implementing our study paradigm with human–robot test subjects.

6 | CONCLUSION

A steadily growing topic within human–robot teaming research is trust and blame. With robots advancing in sophistication and communicative ability, effective cooperation in human–robot teams

will depend on people's social appraisals of their teammate's trustworthiness (Chiou & Lee, 2021). Trust during joint action may be conceptualized as the attitude that a partner (i.e., the trustee) will fulfill the expectations of their role in relation to the trustor (Sheng et al., 2010). Related to these teaming arrangements is the concept of surprise, an emotional or cognitive reaction that often proceeds from unexpected changes in the task environment that can arise during task coordination (Foster & Keane, 2015; Landman et al., 2017). Surprise sparks the need for people to reorient themselves within their social environment, and this includes updating their appraisals of other agents (e.g., teammates) within the system (Loewenstein, 2019). Surprising events may test people's trust in robot partners, and prompt people to blame their partners more often for negative outcomes.

The tendency to trust machines initially can take a severe hit when the machine does not meet expectations, this occurs despite the understanding that machines have limitations in dynamic task environments. However, if the machine is an interactive partner with a higher perceived agency, the negative effects of external factors on trust may not be as readily discernable. The present study sheds light on these dynamics in a joint physical coordination task by testing human dyads as a proxy for future human–robot teams. Counter to prevailing theories of attribution bias, participants surprised by unexpected task environment factors seemed to develop increased trust in their partner and more uncertainty in terms of who to blame for a negative performance outcome. Such findings indicate that as machines become increasingly autonomous, social and environmental dynamics will impact trust development in ways that differ from previous studies on trust in technology.

ACKNOWLEDGMENTS

The authors thank members of the RISE and ADAPT laboratories at ASU for their contributions to the study setup and data collection. This study was partially supported by the National Science Foundation (OIA-1936997) and the Air Force Office of Scientific Research (FA9550-18-1-0067).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Gabriel A. León http://orcid.org/0000-0002-2884-8017
Erin K. Chiou http://orcid.org/0000-0002-7201-8483

REFERENCES

Ajoudani, A., Zanchettin, A. M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., & Khatib, O. (2018). Progress and prospects of the human–robot collaboration. *Autonomous Robots*, *42*(5), 957–975. https://doi.org/10.1007/s10514-017-9677-2

Al-Ani, B., Trainer, E., Redmiles, D., & Simmons, E. (2012). Trust and surprise in distributed teams: Towards an understanding of expectations and adaptations. In R. Vatrapu, V. Evers, Akhilesh Nardi, & M. Maznevski (Eds.), *ICIC'12: Proceedings of the 4th International Conference on Intercultural Collaboration* (pp. 97–106). Association for Computing Machinery. https://doi.org/10.1145/2160881.2160897

Alge, B. J., Wiethoff, C., & Klein, H. J. (2003). When does the medium matter? Knowledge-building experiences and opportunities in decision-making teams. *Organizational Behavior and Human Decision Processes*, *91*(1), 26–37. https://doi.org/10.1016/S0749-5978(02)00524-1

Bies, R. J., Tripp, T. M., & Kramer, R. M. (1997). At the breaking point: Cognitive and social dynamics of revenge in organizations. In R. A. Giacalone & J. Greenberg (Eds.), *Antisocial Behavior in Organizations* (pp. 18–36). Sage Publications.

Braam, I. T. J., van Dormolen, M., & Frings-Dresen, M. H. W. (1996). The work load of warehouse workers in three different working systems. *International Journal of Industrial Ergonomics*, *17*(6), 469–480. https://doi.org/10.1016/0169-8141(95)00008-9

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*(2–3), 171–184. https://doi.org/10.1016/0747-5632(93)90005-D

Campbell, L., Simpson, J. A., Boldry, J. G., & Rubin, H. (2010). Trust, variability in relationship evaluations, and relationship processes. *Journal of Personality and Social Psychology*, *99*(1), 14–31. https://doi.org/10.1037/a0019714

Cassell, J., & Bickmore, T. (2000). External manifestations of trustworthiness in the interface. *Communications of the ACM*, *43*(12), 50–56. https://doi.org/10.1145/355112.355123

Chancey, E. T., Bliss, J. P., Yamani, Y., & Handley, H. A. H. (2017). Trust and the compliance–reliance paradigm: The effects of risk, error bias, and reliability on trust and dependence. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(3), 333–345. https://doi.org/10.1177/0018720816682648

Chiou, E. K., & Lee, J. D. (2021). Trusting automation: Designing for responsivity and resilience. *Human Factors*. Advance online publication. https://doi.org/10.1177/00187208211009995

Cuijpers, R. H., Schie, H. T., Koppen, M., Erlhagen, W., & Bekkering, H. (2006). Goals and means in action observation: A computational approach. *Neural Networks*, *19*(3), 311–322. https://doi.org/10.1016/j.neunet.2006.02.004

Davis, W. L., & Davis, D. E. (1972). Internal–external control and attribution of responsibility for success and failure. *Journal of Personality*, *40*(1), 123–136. https://doi.org/10.1111/j.1467-6494.1972.tb00653.x

De Santis, A., Siciliano, B., De Luca, A., & Bicchi, A. (2008). An atlas of physical human–robot interaction. *Mechanism and Machine Theory*, *43*(3), 253–270. https://doi.org/10.1016/j.mechmachtheory.2007.03.003

Ekman, P., & Rosenberg, E. L. (Eds.). (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Oxford University Press.

Ferrin, D. L., & Dirks, K. T. (2003). The use of rewards to increase and decrease trust: Mediating processes and differential effects. *Organization Science*, *14*(1), 18–31. https://doi.org/10.1287/orsc.14.1.18.12809

Foster, M. I., & Keane, M. T. (2015). Why some surprises are more surprising than others: Surprise as a metacognitive sense of explanatory difficulty. *Cognitive Psychology*, *81*, 74–116. https://doi.org/10.1016/j.cogpsych.2015.08.004

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human–robot collaboration. *2007 International Symposium on Collaborative Technologies and Systems, USA*, 106–114. https://doi.org/10.1109/CTS.2007.4621745

Friedman, B. (1995). "It's the computer's fault": Reasoning about computers as moral agents. In J. Miller, Katz, Mack, & L. Marks (Eds.), *CHI '95: Conference Companion on Human Factors in Computing Systems* (pp. 226–227). Association for Computing Machinery. https://doi.org/10.1145/223355.223537

Friedman, B., Kahn, P. H., & Hagman, J. (2003). Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship. In G. Cockton & P. Korhonen (Eds.), *CHI '03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 273–280). Association for Computing Machinery. https://doi.org/10.1145/642611.642660

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, *101*(1), 75–90. https://doi.org/10.1037/0033-2909.101.1.75

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, *115*(3), 811–846. https://doi.org/10.1162/003355300554926

Glasauer, S., Huber, M., Basili, P., Knoll, A., & Brandt, T. (2010). Interacting in time and space: Investigating human–human and human–robot joint action. *19th International Symposium in Robot and Human Interactive Communication, Italy*, 252–257. https://doi.org/10.1109/ROMAN.2010.5598638

Gomes, J. O., Woods, D. D., Carvalho, P. V. R., Huber, G. J., & Borges, M. R. S. (2009). Resilience and brittleness in the offshore helicopter transportation system: The identification of constraints and sacrifice decisions in pilots' work. *Reliability Engineering & System Safety*, *94*(2), 311–319. https://doi.org/10.1016/j.ress.2008.03.026

Groom, V., Chen, J., Johnson, T., Kara, F. A., & Nass, C. (2010). Critic, compatriot, or chump?: Responses to robot blame attribution. *2010 5th ACM/IEEE International Conference on Human–Robot Interaction (HRI), Japan*, 211–217. https://doi.org/10.1109/HRI.2010.5453192

Groysberg, B., & Abrahams, R. (2006, December 1). Lift outs: How to acquire a high-functioning team. *Harvard Business Review*. https://hbr.org/2006/12/lift-outs-how-to-acquire-a-high-functioning-team

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human–robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Hancock, P. A., Lee, J. D., & Senders, J. W. (2021). Attribution errors by people and intelligent machines. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. Advance online publication. https://doi.org/10.1177/00187208211036323

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: The Almere model. *International Journal of Social Robotics*, *2*(4), 361–375. https://doi.org/10.1007/s12369-010-0068-5

Heider, F. (1958). *The Psychology of Interpersonal Relations*. John Wiley & Sons Inc. https://doi.org/10.1037/10628-000

Hinds, P., Roberts, T., & Jones, H. (2004). Whose job is it anyway? A study of human–robot interaction in a collaborative task. *Human–Computer Interaction*, *19*(1), 151–181.

Holden, R. J. (2009). People or systems? To blame is human. The fix is to engineer. *Professional Safety*, *54*(12), 34–41.

Hollnagel, E. (2013). A tale of two safeties. *Nuclear Safety and Simulation*, *4*(1), 1–9.

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cognition & Emotion*, *20*(2), 242–273. https://doi.org/10.1080/02699930500262878

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: A critique. *Trends in Cognitive Sciences*, *9*(1), 21–25. https://doi.org/10.1016/j.tics.2004.11.003

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*, *14*(1), S103–S109. https://doi.org/10.1006/nimg.2001.0832

Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, *34*(2), 107–117. https://doi.org/10.1037/0003-066X.34.2.107

Jones, E. E., & Nisbett, R. E. (1987). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Lawrence Erlbaum Associates, Inc. https://web.mit.edu/curhan/www/docs/Articles/15341_Readings/Social_Cognition/Jones_&_Nisbett_The_Actor_&_the_Observer_Attribution_pp79-94.pdf

Jung, M. F., Šabanović, S., Eyssel, F., & Fraune, M. (2017). Robots in groups and teams. In C. P. Lee, S. Poltrock, L. Barkhuus, Borges, & W. Kellogg (Eds.), *CSCW '17 Companion: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 401–407). Association for Computing Machinery. https://doi.org/10.1145/3022198.3022659

Kaniarasu, P., & Steinfeld, A. M. (2014). Effects of blame on trust in human–robot interaction. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. 850–885. https://doi.org/10.1109/ROMAN.2014.6926359

Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, *99*(1), 49–65. https://doi.org/10.1016/j.obhdp.2005.07.002

Kim, T., & Hinds, P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. *ROMAN 2006—The 15th IEEE International Symposium on Robot and Human Interactive Communication*, UK, 80–85. https://doi.org/10.1109/ROMAN.2006.314398

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. In W. B. Rouse & K. R. Boff (Eds.), *Organizational Simulation* (pp. 139–184). Wiley. https://doi.org/10.1002/0471739448.ch6

Kochan, J., Breiter, E., & Jentsch, F. (2005). Surprise and unexpectedness in flying: Factors and features. *2005 International Symposium on Aviation Psychology, USA*, 398–403. https://corescholar.libraries.wright.edu/isap_2005/59

Kohn, S. C., de Visser, E. J., Wiese, E., Lee, Y.-C., & Shaw, T. H. (2021). Measurement of trust in automation: A narrative review and reference guide. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.604977

Krosgaard, M. A., Brodt, S. E., & Whitener, E. M. (2002). Trust in the face of conflict: The role of managerial trustworthy behavior and organizational context. *Journal of Applied Psychology*, *87*(2), 312–319. https://doi.org/10.1037/0021-9010.87.2.312

Kuz, S., Petruck, H., Heisterüber, M., Patel, H., Schumann, B., Schlick, C. M., & Binkofski, F. (2015). Mirror neurons and human–robot interaction in assembly cells. *Procedia Manufacturing*, *3*, 402–408. https://doi.org/10.1016/j.promfg.2015.07.187

Landman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Lassiter, G. D., Geers, A. L., Munhall, P. J., Ploutz-Snyder, R. J., & Breitenbecher, D. L. (2002). Illusory causation: Why it occurs. *Psychological Science*, *13*(4), 299–305. https://doi.org/10.1111/j.0956-7976.2002..x

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, K. M., Park, N., & Song, H. (2005). Can a robot be perceived as a developing creature?: Effects of a robot's long-term cognitive developments on its social presence and people's social responses toward it. *Human Communication Research*, *31*(4), 538–563. https://doi.org/10.1111/j.1468-2958.2005.tb00882.x

Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 114–139). Sage Publications. https://doi.org/10.4135/9781452243610

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, *32*(6), 991–1022. https://doi.org/10.1177/0149206306294405

Loewenstein, J. (2019). Surprise, recipes for surprise, and social influence. *Topics in Cognitive Science*, *11*(1), 178–193. https://doi.org/10.1111/tops.12312

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147–186. https://doi.org/10.1080/1047840X.2014.877340

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*, *6*(1), 57–87. https://doi.org/10.1177/1555343411433844

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*(2), 320–339. https://doi.org/10.1111/j.1756-8765.2009.01022.x

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709-734. https://doi.org/10.2307/258792

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24–59.

McDaniel, R. R., Jordan, M. E., & Fleeman, B. F. (2003). Surprise, surprise, surprise! A complexity science view of the unexpected. *Health Care Management Review*, *28*(3), 266–278. https://doi.org/10.1097/00004010-200307000-00008

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(2), 194–210. https://doi.org/10.1518/001872008X288574

Meyer, W.-U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. *Cognition & Emotion*, *5*(4), 295–311. https://doi.org/10.1080/02699939108411042

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*(5), 711–747. https://doi.org/10.1037/0033-2909.130.5.711

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, *82*(2), 213–225. https://doi.org/10.1037/h0076486

Miner, A. S., Bassof, P., & Moorman, C. (2001). Organizational improvisation and learning: A field study. *Administrative Science Quarterly*, *46*(2), 304–337. https://doi.org/10.2307/2667089

Moll Van Charante, E., Cook, R. I., Woods, D. D., Yue, L., & Howie, M. B. (1992). Human–computer interaction in context: Physician interaction with automated intravenous controllers in the heart room. *IFAC Proceedings Volumes*, *25*(9), 263–274. https://doi.org/10.1016/B978-0-08-041900-8.50044-X

Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human–computer interaction. *International Journal of Human–Computer Studies*, *49*(1), 79–94. https://doi.org/10.1006/ijhc.1998.0199

Moore, D. A., Swift, S. A., Sharek, Z. S., & Gino, F. (2010). Correspondence bias in performance evaluation: Why grade inflation works. *Personality and Social Psychology Bulletin*, *36*(6), 843–852. https://doi.org/10.1177/0146167210371316

Morgan, T. (1992). Competence and responsibility in intelligent systems. *Artificial Intelligence Review*, *6*(2), 217–226. https://doi.org/10.1007/BF00150235

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others*. Guilford Press.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man–Machine Studies*, *27*(5–6), 527–539. https://doi.org/10.1016/S0020-7373(87)80013-5

Mutlu, B., & Forlizzi, J. (2008). Robots in organizations: The role of workflow, social, and environmental factors in human–robot interaction. In T. Fong, K. Dautenhahn, M. Scheutz (Eds.), *Proceedings of the 3rd ACM/IEEE International Conference on Human–Robot Interaction* (pp. 287–294). Association for Computing Machinery. https://doi.org/10.1145/1349822.1349860

Onnasch, L., & Hildebrandt, C. L. (2022). Impact of anthropomorphic robot design on trust and attention in industrial human–robot interaction. *ACM Transactions on Human–Robot Interaction*, *11*(1), 1–24. https://doi.org/10.1145/3472224

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, *21*(2), 270–297. https://doi.org/10.1080/08995600902768800

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of Personality and Social Psychology*, *91*(2), 295–315. https://doi.org/10.1037/0022-3514.91.2.295

Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014). Startle and surprise on the flight deck: Similarities, differences, and prevalence. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1047–1051. https://doi.org/10.1177/1541931214581219

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670. https://doi.org/10.1038/35090060

Rochlin, G. I., Porte, T. R. L., & Roberts, K. H. (1987). The self-designing high-reliability organization: Aircraft carrier flight operations at sea. *Naval War College Review*, *40*(4), 76–92.

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion*, *10*(3), 241–278. https://doi.org/10.1080/026999396380240

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology*, *10*, 173–220. https://doi.org/10.1016/S0065-2601(08)60357-3

Sabherwal, R. (1999). The role of trust in outsourced IS development projects. *Communications of the ACM*, *42*(2), 80–86. https://doi.org/10.1145/293411.293485

Sampson, J. P. (1986). Computer technology and counseling psychology: Regression toward the machine? *The Counseling Psychologist*, *14*(4), 567–583. https://doi.org/10.1177/0011000086144006

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human*

Factors: The Journal of the Human Factors and Ergonomics Society, *37*(1), 5–19. https://doi.org/10.1518/001872095779049516

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, behavioral and social dynamics* (pp. 281–308). Springer. https://doi.org/10.1007/978-3-540-74479-5-14

Schützwohl, A., & Borgstedt, K. (2005). The processing of affectively valenced stimuli: The role of surprise. *Cognition & Emotion*, *19*(4), 583–600. https://doi.org/10.1080/02699930441000337

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Sedikides, C., Campbell, W. K., Reeder, G. D., & Elliot, A. J. (1998). The self-serving bias in relational context. *Journal of Personality and Social Psychology*, *74*(2), 378–386. https://doi.org/10.1037/0022-3514.74.2.378

Sharek, Z., Swift, S., Gino, F., & Moore, D. (2010). Not as big as it looks: Attribution errors in the perceptual domain. In M. C. Campbell, J. Inman, & R. Pieters (Eds.), *NA—Advances in Consumer Research* (Vol. 37, pp. 652–653). The harv Association for Consumer Research. http://www.acrwebsite.org/volumes/15445/volumes/v37/NA-37

Sheng, C.-W., Tian, Y.-F., & Chen, M.-C. (2010). Relationships among teamwork behavior, trust, perceived team support, and team commitment. *Social Behavior and Personality: An International Journal*, *38*(10), 1297–1305. https://doi.org/10.2224/sbp.2010.38.10.1297

Stompff, G., Smulders, F., & Henze, L. (2016). Surprises are the benefits: Reframing in multidisciplinary design teams. *Design Studies*, *47*, 187–214. https://doi.org/10.1016/j.destud.2016.09.004

Sundar, S. S., & Nass, C. (2000). Source orientation in human–computer interaction: Programmer, networker, or independent social actor. *Communication Research*, *27*(6), 683–703. https://doi.org/10.1177/009365000027006001

Takayama, L. (2009). Making sense of agentic objects and teleoperation: In-the-moment and reflective perspectives. In M. Scheutz, F. Michaud, P. Hinds, & B. Scassellati (Eds.), *HRI '09: Proceedings of the 4th ACM/IEEE International Conference on Human–Robot Interaction* (pp. 239–340). Association for Computing Machinery. https://doi.org/10.1145/1514095.1514155

Taylor, D. M., & Doria, J. R. (1981). Self-serving and group-serving bias in attribution. *The Journal of Social Psychology*, *113*(2), 201–211. https://doi.org/10.1080/00224545.1981.9924371

Timms, M. (2022, February 9). *Blame culture is toxic. Here's how to stop it*. Harvard Business Review. https://hbr.org/2022/02/blame-culture-is-toxic-heres-how-to-stop-it

Vilaza, G. N., Campos, A. M. C., Haselager, W. F. G., & Vuurpijl, L. (2014). Using games to investigate sense of agency and attribution of responsibility. *Proceedings of SBGames*, *Brazil*, 393–399. https://www.sbgames.org/sbgames2014/papers/culture/full/

Wall, J. A., & Callister, R. R. (1995). Conflict and its management. *Journal of Management*, *21*(3), 515–558. https://doi.org/10.1177/014920639502100306

Walther, J. B., & Bazarova, N. N. (2007). Misattribution in virtual groups: The effects of member distribution on self-serving bias and partner blame. *Human Communication Research*, *33*(1), 1–26. https://doi.org/10.1111/j.1468-2958.2007.00286.x

Wisse, F. (2010). *Effects of adaptive support on team performance by advising human reliance decision making and adaptive automation* [Master Thesis, University of Utrecht]. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cb4a182f760e387b7482f853f504ab2b70f46ca7

Woods, D. D., Sarter, N. B., & Billings, C. (1997). Automaton surprises. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd ed., pp. 1926–1943). Wiley.

Wortham, R. H., Theodorou, A., & Bryson, J. J. (2016). What does the robot think? Transparency as a fundamental design requirement for intelligent systems. *Proceedings of the IJCAI Workshop on Ethics for Artificial Intelligence: International Joint Conference on Artificial Intelligence*., *USA*. https://core.ac.uk/download/pdf/161916101.pdf

You, S., Nie, J., Suh, K., & Sundar, S. S. (2011). When the robot criticizes you…: Self-serving bias in human-robot interaction. In A. Billard, P. Kahn, J. A. Adams, & G. Trafton (Eds.), *HRI '11: Proceedings of the 6th International Conference on Human–Robot Interaction* (pp. 295–296). https://doi.org/10.1145/1957656.1957778

Young, J. E., Hawkins, R., Sharlin, E., & Igarashi, T. (2009). Toward acceptable domestic robots: Applying insights from social psychology. *International Journal of Social Robotics*, *1*(1), 95–108. https://doi.org/10.1007/s12369-008-0006-y

Zaccaro, S. J., Peterson, C., & Walker, S. (1987). Self-serving attributions for individual and group performance. *Social Psychology Quarterly*, *50*(3), 257–263. https://doi.org/10.2307/2786826

How to cite this article: Hsiung, C.-P., León, G. A., Stinson, D., & Chiou, E. K. (2023). Blaming yourself, your partner, or an unexpected event: Attribution biases and trust in a physical coordination task. *Human Factors and Ergonomics in Manufacturing and Service Industries*, *33*, 379–394. https://doi.org/10.1002/hfm.20998