# Attribution Biases and Trust Development in Physical Human-Machine Coordination:

# Blaming Yourself, Your Partner or an Unexpected Event

by

Chi-Ping Hsiung

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2019 by the
Graduate Supervisory Committee:

Erin Chiou, Co-Chair
Nancy Cooke, Co-Chair
Wenlong Zhang

ARIZONA STATE UNIVERSITY

May 2019

# ABSTRACT

Reading partners' actions correctly is essential for successful coordination, but interpretation does not always reflect reality. Attribution biases, such as self-serving and correspondence biases, lead people to misinterpret their partners' actions and falsely assign blame after an unexpected event. These biases thus further influence people's trust in their partners, including machine partners. The increasing capabilities and complexity of machines allow them to work physically with humans. However, their improvements may interfere with the accuracy for people to calibrate trust in machines and their capabilities, which requires an understanding of attribution biases' effect on human-machine coordination. Specifically, the current thesis explores how the development of trust in a partner is influenced by attribution biases and people's assignment of blame for a negative outcome. This study can also suggest how a machine partner should be designed to react to environmental disturbances and report the appropriate level of information about external conditions.

TABLE OF CONTENTS

Page

# LIST OF TABLES

Table Page

# LIST OF FIGURES

Figure Page

# CHAPTER 1

# INTRODUCTION

The robotics and human factor communities have shown growing interest in the concept of machines working as teammates alongside human operators (Schraft, Meyer, Parlitz, & Helms, 2005; Santis, Siciliano, De Luca & Bicchi, 2008; Lien & Verl, 2009; Unhelkar, Siu & Shah, 2014). In light of recent achievement in robotics, machines can now be designed to interact more closely with humans and partner with them to complete a variety of joint physical tasks. Joint physical tasks specifically require the coordination of two or more agents that often demand haptic joint action (Agravante, Cherubini, Bussy & Kheddar, 2013; Cherubini, Passama, Crosnier, Lasnier & Fraisse, 2016; Granados, Yamamoto, 2017; Kucukyilmaz & Demiris, 2018). To achieve successful coordination, it is essential to possess an appropriate level of trust in a machine partner. By guiding a person's reliance, trust can influence human intention and behavior when interacting with machines (robots and automation). Trust thus plays an important role in human-machine coordination (Parasuraman & Riley, 1997; Freedy, DeVisser, Weltman & Coeyman, 2007; Groom & Nass, 2007; Chen & Barnes, 2014; Kaniarasu & Steinfeld, 2014) and should be taken into account in physical coordination as well.

Although there are many studies have already addressed on the indicators that explain people's behaviors in physical human-machine coordination such as interaction force (Jarrassé, Charalambous & BurdetLi, 2012; Mörtl, Lawitzky, Kucukyilmaz, Sezgin, Basdogan &, 2012), few of them consider people's trust in machines. As machines now can be designed to work with people closely as partners, to further explore people's trust in machines during physical coordination it is therefore relevant and necessary in order to achieve successful acceptance and use of machine partners.

## Trust in machine

The development of trust is a dynamic closed-loop process. When people interact with machines (robots and automations), the quality of the interaction with machine influences people's level of trust. This fact is best understood by considering the effect of failures on trust. A human operator's trust in machine declines when severe failures happen and then recovers gradually during the following successful interaction. Further, trust declines with accumulated failures until the operator understands the fault and learns to accommodate it (Itoh, Abe & Tanaka,1999; Lee & Moray, 1992; Lee & See, 2004).

There are multiple factors that can influence an individual's trust in a machine. When people are following the information provided by machines, such as a pilot in the aviation context, there are three elements that influence the trust humans build towards machines and their uses: purpose, process, and performance. The purpose factor is related to the level of automation used, the process factor relates to whether the automated system employed is suitable for the specific task, while the performance factor relates to the system's reliability, predictability, and capability (Lee & See, 2004). Furthermore, people's trust in a machine can be affected by the degree of the perceived transparency and available observability of the automation to the human operator (Verberne, Ham & Midden, 2012). In addition, task complexity, performance saliency, and decision freedom have been suggested to have an impact on the level to which the human operator relies on the automated system (Parasuraman, Molloy & Singh, 1993; Mazney, Reichenbach & Onnasch, 2012; Hoff & Bashir, 2015).

The underlying cognitive schemes that people tend to apply to filter information from other agents can affect the degree of trust of the human operator in machines as well. People tend to rely more readily on the information provided by the machine compared to a human aide, though the content and the delivered method provided both by the automation and the human aid are similar (Dzindolet et al, 2001). One possible explanation for this difference between trust in humans and trust in machine is the cognitive schemas people apply to filter the assessments of aid behaviors. Cognitive schemas are either high expectation assessments in the case of automation or low expectation assessments in the case of humans (Dzindolet, Pierce, Beck & Dawe, 2002). Such filtering of observed aid behavior prompts operators to adopt a particular aid monitoring strategy. For example, whether to be aware of the aid's behavior when interacting with automation; or less sensitive to the errors, making them less noticeable while interacting with a human advisor. This monitoring strategy combines with the primary bases of human trust judgements, which are either performance-linked or knowledge-linked (Lerch, Prietula & Kulik, 1997).

Moreover, a followed-up research concluded that people often exhibit a positivity bias in their trust of novel machine (Dzindolet, Peterson, Pomranky, Pierce & Beck, 2003). This positivity bias could also be a reason that human operators have an unrealistically high expectation of the aid reliability from machine. Therefore, the operator's initial trust in machines is based on faith (Hoff & Bashir, 2015). However, this trust rapidly dissolves following system errors; as relationships with machines progress, dependability and predictability replace faith as the primary basis of trust (Madhavan & Wiegmann, 2007b).

By guiding the human operator's reliance, trust can influence human intention and behavior when coordinating with machines. It is defined as an attitudinal judgment of the extent to which the human operator can rely on the information obtained by automation to achieve their goals, particularly in situations involving risk and uncertainty (Lee & See, 2004; Freedy et al, 2007; Park, Jenkins & Jiang, 2008). Operators who trust a machine more tend to rely more heavily upon it, while individuals with low levels of trust in automation tend to rely less on automation (de Vries, Midden, & Bouwhuis, 2003; Lee & Moray, 1992; Merritt, 2011; Merritt & Ilgen, 2008; Wang, Jamieson, & Hollands, 2009).

The importance of appropriate trust in a machine cannot be emphasized enough. Inappropriate trust in machines can lead to negative repercussions (Dzindolet, Pierce, Beck, Dawe & Anderson, 2001; Visser & Parasuraman, 2011; Chen & Barnes, 2012; Wickens, Hollands, Banbury & Parasuraman, 2015; Robinette, Allen, Howard & Wagner, 2016). For example, a recent study about the investigation of the consequences when people over trust in an autonomous robot in emergency evacuation scenarios (Robinette, Allen, Howard & Wagner, 2016). Therefore, calibrating appropriate levels of trust in a machine based on the exact capacities of the machine is vital for the success of the interaction (Parasuraman & Miller, 2004).

## Trust Calibration and Attribution

*Trust calibration* is defined as the correspondence between a person's trust in the automation and the exact capabilities of the machine (Lee & Moray, 1994). Based on the perceived capability of the machine and the quality of interaction, people determine an estimated calibration for a machine. A well-studied concept in the psychology field that correlates to this process is *attribution*. Attribution is an action when people create a causal explanation for other's or their own behavior based on the perceived information in their social environment. When people coordinate with machines, attribution serves as a tool for people to interpret machines' actions and use the interpretation as a reference of machines' capability, thus, deciding the amount of trust people possess in the machines.

However, attribution is often difficult when people cannot gain much insight into the behavior underlying a system. As machines become more complex, it becomes necessary to have efficient sharing of responsibility between the human and the machines based on their capabilities for a more efficient coordination. It is also important to have a way to clarify the responsibility of an unwanted result. To be more specific, in order to identify reliability of a machine as well as recognize its capability, tracking the sources of errors that occur during the coordination is necessary (Kaniarasu & Steinfeld, 2014). Moreover, error tracking can help people calibrate their trust in machines. A concept closely related to error tracking is *blame*. Error tracking is the process of identifying the cause of an error; blaming is the act of holding the cause of a negative outcome at fault (Parasuraman & Riley, 1997). Deciding who or what to blame is an important part of making sense of complex, difficult situations. In addition, blame means that people can be held accountable (post-hoc) for negative outcomes that were their responsibility (predetermined), and aid in pro-social activity. However, *attribution bias* results when certain factors influence people to attribute blame inaccurately, thus affecting human trust in machines.

There has been some prior work in studying blame in the context of human machine coordination which provides evidence that the attribution bias can be observed when people blame the machine partner for the negative result. For example, Kim and Hinds (2006) observed the application of an autonomous robot that collaborated with a group of nurses in a real-world hospital setting. They found people tend to blame others for errors more than they blame themselves. In addition, when workers noticed inexplicable behavior or errors by the robot, nurses often blamed coworkers for having done something to mess up the robot. Relevantly, Eunil et al (2011) have touched upon the topic of how blame/credit attribution affects the user's trust on the robot. However, their study was focused on how positive or negative feedback would impact user trust in the robot and acceptance of the robot. Kaniarasu and Steinfeld (2014) investigated how the robot assigns blame for an error which affects people's trust in robots. The result indicated that the introduction of blame attribution by the robot lowers user trust in the robot. In addition, users feel positively toward the robot that gives them credit and lack trust in the robot when it degrades them. Although the above research was principally designed to study and explain the factors involved in blame as it relates to negative outcome, researchers actually put little effort into exploring the relationship between people's blame, their trust in a machine and attribution.

People constantly make attributions regarding the cause of their own and others' behaviors; however, rather than operating as objective perceivers, people tend to contribute to perceptual errors that lead to biased interpretations of their social world (Funder, 1987). This phenomenon may be the result of attribution bias, and encompasses a range of related biases. Attribution biases have been shown to direct people's assignment of blame subconsciously. Particularly, there are two common attribution biases that correlate to the development of human trust in machines during coordination. These biases are correspondence bias (Wisse, 2010) and self-serving bias (Madhavan & Wiegmann, 2004).

## Attribution Biases and Human Machine Coordination

*Correspondence bias* is the tendency to judge a person based on their internal characteristics rather than the external situation they might be facing (Jones, 1979; Gilbert & Malone, 1995). That means when two agents work together, if environmental instability occurs and influences their partner's behavior, people tend to overestimate the role of dispositional factors and underestimate the role of environmental instability. Relatedly, correspondence bias is identified as a potential explanation for inappropriate trust calibration (Wisse, 2010). Muir (1987) argued that people are more likely to attribute a perceived unpredictability of a machine to the machine's properties than to environmental instability, even when the environmental instability is the main cause of the machine's unpredictable behavior. Moreover, when a person assigns responsibility with correspondence bias to a machine partner, he/she will tend to underestimate the machine's predictability and dependability. Because predictability and dependability are critical factors that affect people's trust (Lee & Moray, 1992; Muir & Moray, 1996), human trust in the machine will decrease. Thus, this phenomenon should be taken into account when designers and researchers design machines to avoid misuse and disuse of machine.

However, the possibility is raised that the correspondence bias, as demonstrated previously, might simply be a problem of incomplete information. In Ross et al.'s (1977) experiment, as in many studies that demonstrate correspondence bias, it was difficult for individual participants to precisely determine the strength of the situation. To confirm the robustness of the correspondence bias, recent research suggests that correspondence bias can persist even when information about both behavior and situation are known with equal clarity and are presented in the same format and modality (Moore et al, 2010).

Subsequently, *self-serving bias* refers to the tendency for people to blame negative outcomes on external factors, while giving credit to themselves for positive outcomes (Davis & Davis, 1972; Millar & Ross, 1975). This bias can be observed in both individual activities such as one-on-one sports which clearly define a winner (De Michele, Gansneder & Solomon, 1998) and group activities such as team sports in which the outcome would be distributed among all team members (Lau & Russell, 1980). Besides, previous research indicated that in social interaction, partners' actions are more salient and provide more pertinent cues than the environment (Jones & Nisbett, 1971). Therefore, when people coordinate with others, they are more likely to attribute the negative outcome to their partner compared to themselves or environmental factors (Walther & Bazarova, 2007). In addition, when people coordinate with others as a group, the effect of self-serving bias might be mitigated (Zaccaro, Peterson & Walker, 1987).

Interestingly, previous research has found that in the context of coordination, people tend to blame technology for mistakes and errors while also exhibiting reluctance to credit positive outcomes to their non-human partners (Sampson 1986, Morgan 1992, Friedman 1995), or even to anthropomorphic agents. Moon and Nass (1998) found that if participants were coordinating with a computer assistant whose personality is dissimilar to theirs, they will tend to exhibit self-serving bias, especially in the case of a negative outcome. Similarly, self-serving bias was observed when people interact with robots. You, Nei, Suh and Sundar (2011) found that when a robot was put in the role of instructor and made different types of verbal evaluations of participants' performance, participants tended to dismiss criticism from the robot and attributed blame to the robot, while claiming credit for themselves when their performance was rated positively. Relatedly, self-serving bias was observed in the scenario in which control is shared between participants and machines. Vilaza, Campos, Haselager and Louis (2014) designed a computer game which requires both an AI (artificial intelligence) and a participant to control the direction of a ball to avoid obstacles and collect the target item together. Their findings indicated that participants were shown to blame the AI when they lost a game, whereas they took credit when they won a game. These studies imply that self-serving bias can be observed in the context of human-machine coordination.

Additionally, although the self-serving bias can be observed in varied areas, the underlying cause of the self-serving bias still remains controversial (Shepperd, Malone & Sweeny, 2008). One popular explanation is that the self-serving is associated with the self-protective bias. That is, when people receive negative evaluations, they tend to reject the criticism, while those receiving positive evaluations tend to accept the praise to enhance their self-esteem (Swann & Schroeder, 1995; Sedikides, Campbell, Reeder & Ellio, 1998).

### Attribution Biases and Trust in Machine

Similar to human relationships, in coordinative environments, where two or more entities work together to accomplish the task at hand, these two attribution biases may direct the way people interact with their partners. To be more specific, the development of interpersonal trust can be viewed as an attribution process. For example, an individual may develop beliefs about another person's trustworthiness based on whether the person's behavior is judged to be caused by internal versus external factors. (Krosgaard, Brodt & Whitener, 2002). Also, attribution biases are found to relate to trust in partners. Ferrin and Dirks (2003) manipulated the initial trust levels (high or low) by indicating if their partner shared all relevant information and the accuracy of shared information. Besides initial trust level, three types of reward structures were manipulated. Each pair of participants was assigned to experience evaluation criteria that were either based on the performance of their dyad (cooperative structure), their partner's performance (competitive structure), or half of their dyad and half of their partner's performance (mixed reward structure). Also, each pair of participants was informed that the highest-scoring participants would be included in a lottery to win $75. Their results suggested that the reward structure influences trust, more importantly, people's attribution was able to provide a useful framework for understanding the complex, diverse, and multiple routes through which trust may develop.

Consequently, one pressing question is if the findings on people coordinating could be transferred to human-machine coordination; can people's attribution biases affect a machine's perceived trustworthiness in the same way?

If people interpret the motivation and reason underlying their machine partner's actions incorrectly, their trust and belief in their partner might be built erroneously. With the concept of human and machine coordination being embraced in the near future, the relationship between attribution biases and human trust in machine should be explored in more depth to achieve appropriate trust in machine partners.

Therefore, our present work sets out to understand the assignment of blame to different participating entities involved in the smallest possible unit of physical human-machine coordination (a person, a machine counterpart, and a task environment) and how it impacts human trust in a machine. More importantly, we examine the role of how attribution biases act in the development of human trust in the context of human machine joint activity. We asked:

1. In the context of an unexpected environmental event that causes a partner to behave unpredictably, which results in an unwanted outcome, to which entities will people assign blame for the negative outcome: themselves, their partner, or the environmental event?
2. Following research question 1, can attribution biases be observed in the context of physical coordination?
3. Specifically, following research question 1 and 2, is there any significant difference in the assignment of blame between physical coordination with a human partner and a machine partner?
4. To what extent do attribution biases influence human trust in their partner in the context of physical coordination?
5. Following research question 4, does trust develop differently when coordinating with a human partner compared to a machine partner?

# CHAPTER 2

## METHOD

We seek to gain insight into how human trust in a machine develops after experiencing an unexpected event during coordination. Also, we explored the potential relationship between attribution biases and trust. From previous findings, we know that with the engagement of self-serving bias, people tend to blame a negative outcome to their partner and are less likely to blame themselves. Furthermore, with the introduction of correspondence bias, people tend to consider that their partner's behavior is due to their internal characteristics rather than external factors that the partner might be facing. Therefore, we can conclude that during a coordination with an unexpected event, if the outcome is negative, people would tend to blame the negative outcome to their partner more when compared with the same coordination without the unexpected because people would consider their partner's internal characteristics indirectly affect the result. The hypotheses are formally stated as follows:

**H1: Effect of self-serving bias on attribution of blame,** *participants will be more likely to blame their partner for the negative outcome, rather than to blame themselves (or the unexpected event).*

**H2: Effect of self-serving bias on trust,** *if participants blame the negative outcome on the human partner, rather than themselves (or the unexpected event), participants' degree of trust in their partner will be significantly lower than their initial trust.*

**H3: Effect of the correspondence bias on attribution of blame,** *with the introduction of an unexpected event, participants' degree of blame for a negative outcome on their partner will be significantly higher than the case without an unexpected event.*

**H4: Effect of the correspondence bias on trust,** *with the introduction of an unexpected event, participants' degree of trust in their partner will be significantly lower compared to the case without an unexpected event.*

### Experimental design

To test our hypotheses, we applied a between-subject experimental design with self-reported measures following a joint physical coordination task. The present study randomly assigned participants to coordinate with an unfamiliar human partner and complete one of two conditions – a baseline condition, and a surprise condition that involved a designated unexpected event during a coordinated transportation task in order to investigate the effect of correspondence bias. Self-reported trust in partners, surprise level of an unexpected event, facial expression in response to surprise, and attribution of blame were the dependent variables that were measured.

A transportation task was designed to demonstrate joint physical coordination in a laboratory setting for this study. That is, participants were asked to lift an object with their partner from the designated area on the ground to the assigned table, and then lift the object back to the area. The box, which contained three bricks with 5.75 lbs and a cup with 200ml of water, was used as the object for participants to transport during the coordination task. During the transportation process, participants were asked not only to maintain the stability of their movement to prevent the water spill, but also complete the task as soon as possible.

Also, all pairs of participants were informed that the completion time and the water they keep in the cup during the transportation task would be measured as their performance, and the group with the best performance among all the pairs of participants received a $30 Starbucks gift card. This task would repeat five times in total including one practice trial. To create a standardized environmental instability, we introduced a warning tone as an unexpected event in our study. Participants were told beforehand that during the warning tone, they have to stop moving and stay still, and then, when the warning tone ends, they can continue their previous actions.

Our coordinated transportation task can be seen as a "microworld" study; a microworld is a simplified version of a real system in which the essential elements are retained and the complexities eliminated to make experimental control possible (Brehmer & Dorner, 1993; Lee & See, 2004). The elements we manipulated in the design of the coordinated transportation include motivation, familiarity, and competing demand.

The experimental setting for motivation in this study was derived from the study conducted by Ferrin and Dirks (2003). In the present study, the cooperative reward structure and competitive structure were applied. Participants were informed that their performance will be based on their dyad, also, they had to compete with other pairs of participants to win the monetary prize. Subsequently, in order to bring about participants' sense of blame, at the end of the experiment, participants were told that based on their performance, they did not win the gift card. The intention of using a sense of competition was to create a scenario of failure, without causing actual harm, in which it could be observed how the participants reacted to it.

Second, for familiarity, the joint transportation task is prevalent in our daily life. Most of our participants are already familiar with the nature of the joint transportation task as well as the disturbance that may happen during the task, so that they are likely to learn the task with relatively little training.

Last but not least, people often face competing demands on their time and cognitive resources (Merritt & Ilgen, 2008). To be more specific, in this study, participants have to balance the competing demands of speed and accuracy. To achieve this, participants were told that they had to complete the task as soon as possible and also keep stable to prevent water from spilling in order to simulate the competing demands people might experience in everyday life.

**Participants**

Twenty-six participants (10 females, 16 males) from Arizona State University participated this study. Nineteen out of the twenty-six participants were recruited from an online course credit management system, while the remaining were recruited either via paper flyers or in-person recruitment. All participants reported that they had no prior experience working with manipulator robots, were able to comfortably lift, were not familiar with the other participant with whom they would be coordinating on the designed task, were able to carry 10 lbs with their dominant arm, and were comfortable communicating in English. All participants were required to be at least 18 years old. To aid in study recruitment, participants who were recruited from the online course credit management system would receive one research credit after completing the experiment.

## Equipment and Materials

A motion capture system (Optitrack, Natural Point, Corvallis, OR, USA) was set up to capture the motion behaviors of participants. The physical setup of the equipment relative to participants is shown in Fig. 1 and Fig. 2. Instructions for the task were delivered by researchers using a Powerpoint presentation and a script, to explain the task and spatial stimuli in the task environment. A semitransparent plastic box, which contained three bricks with 5.75 lbs and a cup with 200ml of water, was used as the moving object for the coordinated transportation task.

The task environment included a table, and two blue squares. One marked on one side of the table, and one marked on the ground, both indicating where the box should be placed. Also, another two blue squares on each side of the box contain either number one or number two. Different numbers were used to differentiate the standpoints of the participants. This visual representation of the task was purely to aid participants in task completion and to provide a more controlled setting for communicating each task to participants. Finally, surrounding the area were motion sensors and a desk with a computer for recording data (See Fig 3).

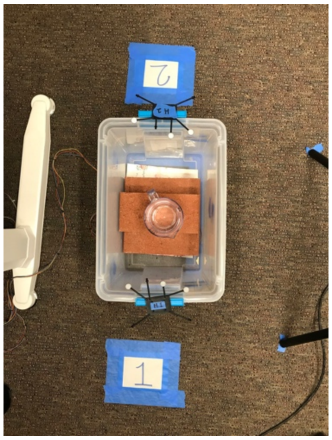

Figure 1. Experiment setup

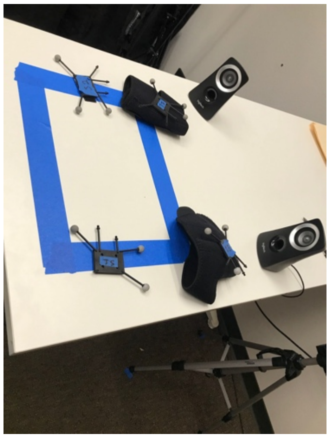

Figure 2. Motion capture devices and speakers

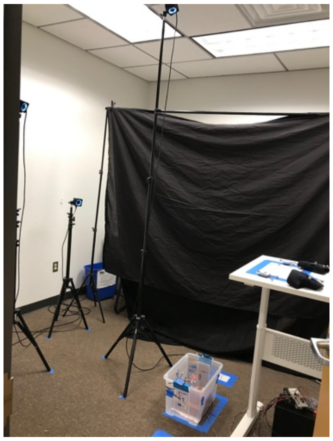

Figure 3. Experiment surrounding

*Unexpected event.* For creating an unexpected event to understand the effect of correspondence bias, we followed the principles about the prior knowledge of the unexpected event made by Kochan, Breiter, and Jentsch (2004). In the study, a 250 Hz tone was played continually for four seconds during the surprise condition at trial two and trial four via Logitech Z313 Speaker System.

Surprise, instead of startle, was implemented in this study in order to minimize and avoid the harm that participants might experience as well as potential uncontrollable situations which might occur during the experiment, the intensity stimuli of an unexpected event can cause different reactions of individuals, such as surprise and startle. 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, Niepel, Rudolph, & Schützwohl, 1991; Schützwohl & Borgstedt, 2005). Unlike startle, which always occurs as a response to the presence of a sudden, high-intensity stimulus surprise can be elicited by an unexpected stimulus or by the unexpected absence of a stimulus (Rivera et al, 2014). In our study, participants were given knowledge in training about how to respond to the tone, which makes the warning tone a surprise, not a startle.

## Measures

*Trust measurements: Muir trust scale.* Based on the findings by Merritt and Llgen (2008), we expect to observe that there are different constructs of trust in the questionnaire results after a short period of interaction. In this study, subjective ratings of trust in automation were obtained on a 5-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree), built after scales used by Merritt and Llgen (2008). One item assesses overall trustworthiness of a partner, and the other four items each relate to the trust-related factors in Muir's (1987) theory.

The exact questionnaire was conducted a couple of times during the process of experiment. The first questionnaire was conducted right after the training slides were introduced by researchers to measure participants' initial trust in their human partner. The second questionnaire was conducted after all the tasks in the experiment were completed. Table 1 shows the Cronbach's alpha for each scale. See Table 2 (Appendix) for a summary of the trust responses in each condition.

Table 1
*Cronbach's Alpha Values of the Muir's Trust Scale of the Different Conditions*

| Baseline | | Surprise | |
|---|---|---|---|
| Initial<br>(*n* = 12) | Post<br>(*n* = 12) | Initial<br>(*n* = 14) | Post<br>(*n* = 14) |
| .781 | .859 | .740 | .825 |

*Note.* Cronbach's Alpha is a measure of the reliability of the scale as a whole. Alpha ranges from zero to 1.0 (highest).

*Attribution measurement.* To access participants' attribution of blame, a categorical questionnaire was used in this study. The constructs of people's responsibility and attribution of blame in a human partner or a machine in the context of coordination has been studied in human-computer interaction (HCI) (Moon& Nass, 1998; Moon, 2003) and recent human-robot interaction (HRI) (Kim & Hinds, 2006; Groom et al, 2010; Kaniarasu & Steinfeld, 2014). In these studies, categorical approach has been used by most of them. This study modified the scale used by Kim and Hinds (2006) to replace the word "other" and “robot” with “partner” and “the warning tone” according to our hypotheses. Also, for this study, we only attempt to understand the effect of blame on human trust, hence, the final outcome would always be negative. Therefore, this study only asks participants their assignment of blame and the level of responsibility for the negative outcome, rather than having the participant attribute both the credit of a successful outcome and blame of an unsuccessful outcome.

The same questionnaire of attribution of an unsuccessful result was implemented in both surprise and baseline conditions. All questions were answered on a 7-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). For each entity for blaming (self, partner, or the warning tone), participants were asked two questions and the final scores were the average scores of the two values. Table 3 shows the Cronbach's alpha for each scale. Table 4 (Appendix) displays the summary of the attribution of blame responses in each condition. This questionnaire was administered through Qualtrics, an online survey tool.

Table 3
*Cronbach’s Alpha Value for the Dependent Variables*

| Scales | Cronbach’s α | |
|---|---|---|
| | Baseline<br>(*n* = 12) | Surprise<br>(*n* = 14) |
| **Attribution of blame to self**<br>- I was responsible for the unsuccessful result<br>- I was to blame for the unsuccessful result | .835 | .986 |
| **Attribution of blame to the partner**<br>- My partner was to blame for the unsuccessful result<br>- My partner was responsible for the unsuccessful result | .869 | .935 |
| **Attribution of blame to the warning tone**<br>- The warning tone was to blame for the unsuccessful result<br>- The warning tone was responsible for the unsuccessful result | .000 | .989 |

*Note.* Cronbach's Alpha is a measure of the reliability of the scale as a whole. Alpha ranges from zero to 1.0 (highest).

*Surprise measurement.* Surprise, not startle, can be measured by both subjective self-report and behavioral methods. In this study, we applied a 7-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree) as self-report measurement to assess the subjective surprise levels participants experienced (Reisenzein et al., 2006). Any participant who reported five or higher on the self-report surprise scale was considered surprised. The self-report surprise scale was used after the warning tone was played. As well, the facial expression checklist for surprise was used to serve as an objective reference of whether or not participants were surprised. Participants' facial expressions were documented by researchers by hand while the warning tone was playing. Any participant who showed at least one facial expression was considered surprised (Ekman & Rosenberg, 1997).

*Participant demographics.* To address potential confounds in explaining the relationship between trust and attribution of blame, demographic measures of age, height, weight, if participants speak English natively, self-identified multi-tasking tendency, and whether or not they were using their dominant hand during the task were included in the analysis. Table 5 provides summative descriptive statistics of demographics for baseline and surprise conditions.

Table 5
Descriptive Statistics of Demographics Information of Participant for Each Condition

| | Baseline<br>(*n* = 12) | | Surprise<br>(*n* = 14) | |
|---|---|---|---|---|
| Factor | *M* | *SD* | *M* | *SD* |
| Height (inches) | 66.67 | 4.186 | 67.71 | 4.746 |
| Factor | | | | |
| Age | | | | |
| 18 – 23 | 92% | | 57% | |
| 23 – 28 | 8% | | 43% | |
| Native Speaker of English | | | | |
| Yes | 9 | | 12 | |
| No | 3 | | 2 | |
| Multitasker | | | | |
| Yes | 11 | | 12 | |
| No | 1 | | 2 | |
| Dominant Hand Usage | | | | |
| Yes | 12 | | 11 | |
| No | 0 | | 3 | |
| Weight (lbs.) | | | | |
| 120 – 160 | 75% | | 29% | |
| 160 – 200 | 25% | | 57% | |
| 200 – 240 | 0% | | 14% | |

*Note.* For continuous and ordinal data, we report the mean and standard deviation, categorical data we report the frequency and percentage.

## Procedure

Upon both participants arrival and greeting, two of the researchers provided each participant with a brief overview of the study and asked him/her to read and sign an informed consent form, fill out a demographic survey, and an initial trust questionnaire. Then, the researchers explained the task to the participant and took him/her through a training session to familiarize the participant with the transportation task, the monetary prize, and instruction participants have to follow during the warning tone. After instructing participants to locate to the designated spot, they would be asked to go through a training trial.

Once confirmed that the participants have no further questions about the experimental task, the participants then performed the experimental task described in the experimental design section four times. In the condition with an unexpected event, a warning tone would be played during trial two and trial four. While the warning tone was played, researchers would document the participants' facial expressions for surprise. Once the warning tone was gone, participants could continue their previous actions. At the end of trials two and four, each participant would be asked how surprised they were. After completing the four trials, the participants would be informed that based on their performance, they did not win the prize. Subsequently, the participants were asked to fill out a post-trust questionnaire. As the final step, we debriefed the participants of the nature of the study. The overall procedure took less than one hour. The motion data and completion time were recorded by the researcher who sat behind the black curtain on the left side of the participant throughout the experiment. The facial expression checklist for the surprised reaction caused by the warning tone was recorded by the researchers standing next to each participant.

### Analysis

Initially, the collected data was analyzed by the Shapiro-Wilk test for normal distribution and Levene's statistic for homogeneity of variance. The data of baseline condition violated homogeneity of variance. Therefore, nonparametric statistics was used for the baseline condition. To interpret the within-subject effects of the different entities participants blamed for the negative outcome, Friedman two-way analysis of variance test with Wilcoxon signed-rank tests were used for the baseline condition and a one-way analysis of variance was applied for the surprise conditions. Further, the difference between initial trust and post-task trust was revealed using paired samples t-test. Following this, to investigate the between-subject effects of the introduction of the warning tone on attribution of blame and trust in partner, two independent sample t-test were executed. Finally, we used Pearson's correlation test and linear regression to further explore the presence of a relationship between our measures within each condition, including initial and post-task trust, and the attribution of blame. To achieve that, we averaged the values of each item in trust questionnaires by participants. Next, we calculated the difference between the initial and post-task trust scores. For the results of the blame questionnaire, the scores of each item were standardized with the scores of partner-blame as a baseline. The higher standardized scores represent the participant assigning more blame to their partner. Lastly, we calculated the difference using the standardized score and saw how they differ between self-blame and warning tone-blame. In order to address potential confounds in explaining the relationship between trust and attribution of blame, participants' demographic information was analyzed by Spearman's correlation with initial trust, post-task trust and the standardized scores of the attribution of blame within each condition. We used an alpha level of .05 for all statistical tests except Wilcoxon signed-rank test. Data were evaluated using SPSS.

In addition, the usage of motion capture devices was for recording participants' motion patterns for a different study about developing an algorithm for robots to collaborate with humans and thus will not be included in this study.

# CHAPTER 3

# RESULT

Twenty-six participants were included in this study. Among them, six pairs of participants were in the baseline condition and the other seven pairs of participants were in the surprise condition. All of the participants were assigned to complete the physical coordination task with another unfamiliar human partner. This preliminary analysis served as a reference which allowed us to gain insight into the further human-machine coordination.

Initially, the collected data were analyzed by the Shapiro-Wilk test for normal distribution. In the baseline condition, blame scores obtained in self-blame were: *df*(10*)* = 0.909*,* *p* = .272, in partner-blame, *df*(10) = 0.871, *p* = .102; and in warning tone-blame: *df*(10) = 0.366*,* *p* < *.01*. The result disclosed the data of warning tone-blame violated the normal distribution. Further, the Levene's statistic for equality of variances indicated a significant difference (*F*(2, 27) = 8.552, *p* = .01) in baseline condition suggesting there was a violation of homogeneity of variance. Therefore, nonparametric statistics was used.

Likewise, for the surprise condition, the Shapiro-Wilk test for blame scores obtained in self-blame were: *df*(12) = 0.926, *p* = .342, in partner-blame, *df*(12) = 0.878, *p* = .082, and in warning tone-blame, *df*(12) = 0.894, *p* = .131. The result indicated that the data were normally distributed. Next, the Levene's statistic for equality of variances indicated no significant difference (*F*(2, 32) = 0.75, *p* = .48) in surprise condition suggesting there was no violation of homogeneity of variance. Therefore, parametric analysis was chosen.

### Effects of Self-Serving Bias

In hypothesis one, we argued that the engagement of self-serving bias would lead participants to make more attributions of blame to their partner and less to the participants themselves or the involvement of the warning tone. Friedman two-way analysis of variance test was executed to explore if there was a statistically significant difference between the self-blame, partner-blame scores, and the warning tone-blame. The result revealed that there was a significant difference ($\chi^2(2)$ =10.47*,* $p$ = .005) between the three blamed entities. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. Median (IQR) for the self-blame, the partner-blame and the warning tone-blame scores were 2.75 (1.88 to 5), 2.25 (1 to 3.25) and 1 (1 to 1), respectively. There were no significant differences between the self-blame and the partner-blame scores ($z$ = -1.38, $p$ = .168) or between the partner-blame and the warning tone-blame scores ($z$ = -2.20, $p$ = .028). On the other hand, there was a statistically significant difference in the self-blame score versus the warning tone-blame score ($z$ = -2.53, $p$ = .011). Further, Kendall's effect size value ($W$ = 0.80) suggested a strong practical significance. Table 6 shows the descriptive statistics and results of Friedman two-way analysis of variance test and Wilcoxon signed-rank test.

Table 6
Descriptive Statistics and Results of Friedman Two-Way Analysis of Variance Test and Wilcoxon Signed-Rank Test of the Different Responses of the Attribution of Blame Questionnaire in the Baseline Condition

| | *N* | *M* | *SD* | Percentiles | | | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| | | | | 25th | 50th<br>(Median) | 75th | | |
| Self-blame | 10 | 3.25 | 1.86 | 1.88 | 2.75 | 5.00 | -1.37 | .168 |
| Partner-blame | 10 | 2.30 | 1.40 | 1.00 | 2.25 | 3.25 | -2.53* | .011 |
| Warning tone-blame | 10 | 1.05 | 0.16 | 1.00 | 1.00 | 1.00 | -2.20 | .028 |

*Note.* * Wilcoxon signed-rank test is significant at the 0.017 level (2-tailed).

For the surprise condition, a one-way analysis of variance indicated that there was no significant difference ($F(2,22) = 0.194$, $p = .83$) on the entities (the participants themselves, their partner, or the warning tone) that participants blamed for the negative outcome. Therefore, we concluded that the self-serving bias was not able to be observed in the present study setting. Table 7 presents the descriptive statistics of the two conditions. Table 8 shows the result of one-way analysis of variance test. Figures 4 and 5 present the means for attribution of blame questionnaires responses (self-blame, partner-blame and warning tone-blame) with 95% confidence intervals for each condition.

Table 7

Descriptive Statistics of the Result of Attribution of Blame Questionnaire Responses in the Surprise Condition

| | *N* | *Mean* | *SD* | *SE* | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| Self-blame | 11 | 3.46 | 1.56 | .47 | 2.41 | 4.50 |
| Partner-blame | 12 | 2.83 | 1.74 | .50 | 1.73 | 3.94 |
| Warning tone-blame | 12 | 3.17 | 2.03 | .59 | 1.88 | 4.45 |

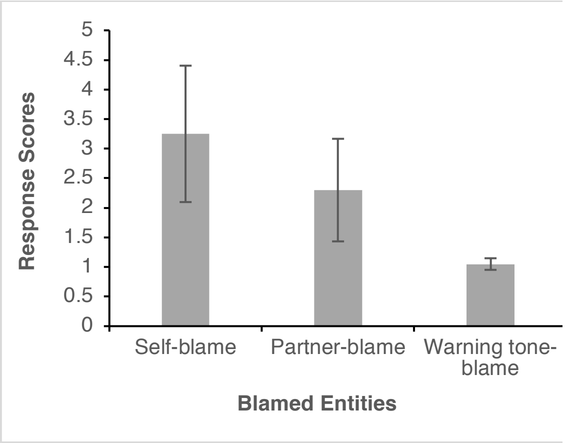

**Figure 4**, Mean response scores (range from 1 to 7) for attribution of blame questionnaire in the baseline condition (*N* = 12). Error bars denote 95% confidence intervals.

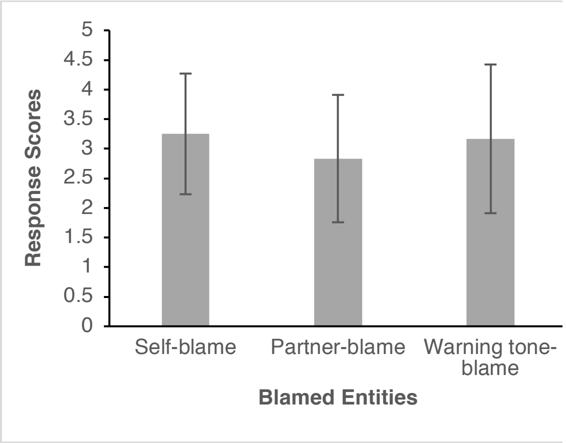

**Figure 5**, Mean response scores (range from 1 to 7) for attribution of blame questionnaire in the surprise condition (*N* = 14). Error bars denote 95% confidence intervals.

Table 8
One-Way Analysis of Variance of the Result of Attribution of Blame Questionnaire Responses in the Surprise Condition

| Source | *df* | *SS* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Between Groups | 2 | 2.23 | 1.11 | .347 | .709 |
| Within Groups | 32 | 102.56 | 3.21 | | |
| Total | 34 | 104.79 | | | |

On average, in the baseline conditions, participants tend to blame themselves (*M* = 3.25, *SD* = 1.86) for the negative result compared to when they blame their partner (*M* = 2.30, *SD* = 1.40). On the other hand, participants in the surprise condition are more likely to blame themselves (*M* = 3.46, *SD* = 0.16) compared to when they blame the warning tone (*M* = 3.17, *SD* = 2.03) and their partner (*M*= 2.83, *SD* = 1.74). However, these differences were not statistically significant.

In hypothesis two, we argued that participants' trust would decrease with the engagement of the self-serving bias. However, even the self-serving bias cannot be observed in the present study, our result suggesting the collected data were opposite to our hypothesis in both conditions. That is, participants trust in their partner more after completing the coordination task. In the baseline condition, the participants' post-task trust score is higher compared to their initial trust score, but the difference was not significant (*t*(11) = -1.91, *p* = .08). In the surprise condition, the result of pair sample t-test revealed a significant difference (*t*(13) = -2.33, *p* = .037; *d* = 0.62) between participants' post-task trust and the initial trust in their partner. Cohen's *d* suggested that the effect size of this analysis was found to have moderate effect (*d* = 0.60). Table 9 (Appendix) presents the descriptive statistics of the two conditions. Figures 6 and 7 present the means for initial and post-task trust scales responses with 95% confidence intervals for each condition.

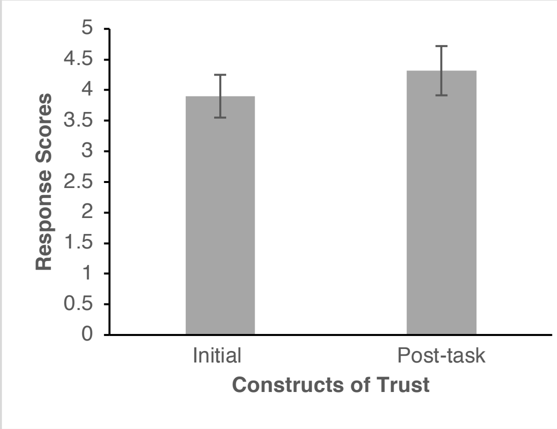

**Figure 6**, Mean response scores (range from 1 to 5) for initial and post-task trust scales in the baseline condition. Error bars denote 95% confidence intervals.

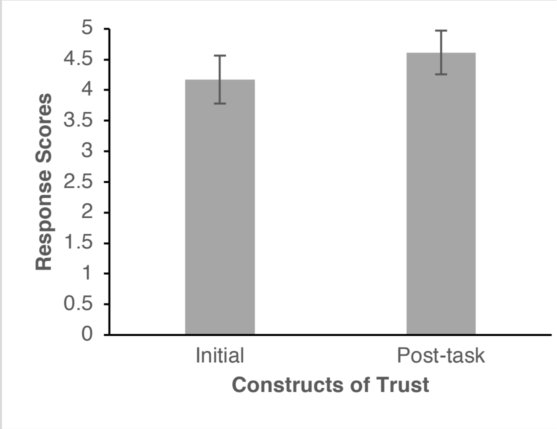

**Figure 7**, Mean response scores (range from 1 to 5) for initial and post-task trust scales in the surprise condition. Error bars denote 95% confidence intervals.

## Effects of Correspondence Bias

In hypothesis three, we argued that with the introduction of the warning tone, participants in the surprise condition would be more likely to blame their partner compare to baseline condition. However, the results indicated no support for this hypothesis. There was little difference (*t*(20) = 0.001, *p* = .999) between in the standardized scores of blame attribution to the partner in the baseline condition (*M* = 0.00, *SD* = 1.72) as compared with the surprise condition (*M* = 0.0008, *SD* = 2.11). The hypothesis four predicted that with the introduction of an unexpected event, participants' degree of trust in their partner will be significantly lower compared to the case without an unexpected event. That is, the decrease of trust in the partners for the surprise condition (*M* = 0.417, *SD* = 0.76) should be larger than the baseline condition (*M* = 0.443, *SD* = 0.71). However, there was no significant difference found in the result of independent t-test (*t*(24) = -0.91, *p* = .928). These results suggested correspondence bias cannot be observed in the present experimental condition. Tables 10 and 11 present the results of an independent sample t-test for both conditions. Figure 8 shows the means for the differences of the average initial and post-task trust scores in each condition.

Table 10
Results of an Independent Sample T-Test Between the Conditions for the Attribution of Blame

| | Condition | *N* | *Mean* | *SD* | *SE* | Paired difference of mean | *t* |
|---|---|---|---|---|---|---|---|
| Attribution of blame | Baseline | 10 | .000 | 1.72 | 0.54 | -.00083 | -.001 |
| | Surprise | 12 | .001 | 2.11 | 0.61 | | |

Table 11
Results of an Independent Sample T-Test Between the Conditions for the Trust Difference

| | Condition | *N* | *Mean* | *SD* | *SE* | Paired difference of mean | *t* |
|---|---|---|---|---|---|---|---|
| Trust difference | Baseline | 12 | .417 | .7554 | .2181 | -.0262 | -.091 |
| | Surprise | 14 | .443 | .7111 | .1901 | | |

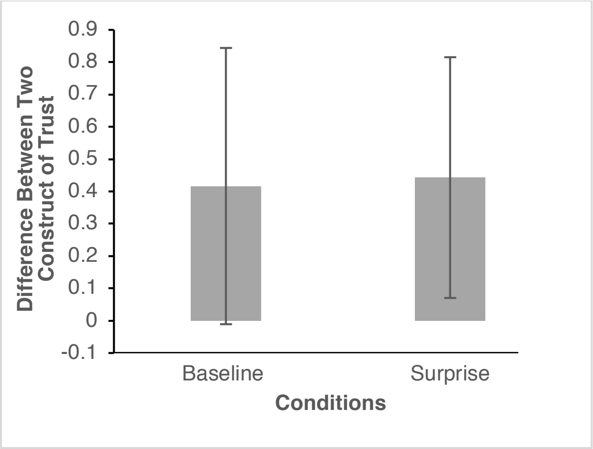

**Figure 8**, Mean scores for the differences of the average initial and post-task trust scores in each condition. Error bars denote 95% confidence intervals.

Finally, Pearson's correlation indicated that in the baseline condition, attribution of blame and the post-task trust in the partner were strongly negatively correlated (*r*(13) = -.636, *p* = .048). However, there was no significant correlation found between initial trust or the development of trust with attribution of blame (see table 12 for the result of correlation). To further examine the correlations, linear regressions of the mean values between the post-task trust and the attribution of blame showing significant differences. The results indicated that the entities people blame for the negative outcome explained 40.4 % of the variance ($R^2$ = .404, *F*(1,8) = 5.427, *p* = .048) in the post-task trust. Table 13 presents the result of linear regression. The significant predictor of the post-task trust was the attribution of blame in their partner. On the other hand, in the surprise condition, Pearson's correlation test revealed that there was no significant result found between each variable (see table 14 for the result of correlation).

Table 12
The Result of Correlations Between the Initial trust, Post-Task Trust and Attribution of Blame for the Baseline Condition

| | 1 | 2 | 3 |
|---|---|---|---|
| 1. Initial trust | – | | |
| 2. Post-task trust | .360 | – | |
| 3. Attribution of blame | -.565 | -.636* | – |

*Note.* * Correlation is significant at the 0.05 level (2-tailed).

Table 13
Linear Regression of Post-Task Trust Predicted by the Attribution of Blame for the Baseline Condition

| Source | | | | | |
|---|---|---|---|---|---|
| | *B* | *SE B* | *β* | *t* | *p* |
| Attribution of blame | -.286 | .123 | -.636 | -2.330* | .048 |

*Note.* * Linear regression is significant at the 0.05 level (2-tailed).

Table 14
The Result of Correlations Between the Initial Trust, Post-Task Trust and Attribution of Blame for the Surprise Condition

| | 1 | 2 | 3 |
|---|---|---|---|
| 1. Initial trust | – | | |
| 2. Post-task trust | .508 | – | |
| 3. Attribution of blame | .195 | .147 | – |

Further, Spearman's correlation indicated that in the baseline condition, participants' age and their self-report recognition of multitasker was strongly negative correlated (*r*(12) = -1.00, *p* < .01). Also, whether or not the participants are native speakers of English was strongly negative correlated with self-identified as a multitasker (*r*(12) = -1.00, *p* < .01). See table 15 (Appendix) for the result of correlation.

In the surprise condition, participants' heights were found to correlate with initial trust (*r* (14) = .619, *p* = .02) and post-task trust (*r*(14) = .659, *p* = .01) with large effects. Also, the result revealed that participants' dominant hand usage strongly correlated to participants' age (*r*(14) = -.603, *p* = .02), their self-report recognition of multitasker (*r*(14)

= .782, *p* = .001) and if they are native speaker of English (*r*(14) = .782, *p* = .001). Besides, the result indicated that participants' weights are strongly correlated to participants' age(*r*(14) = .644, *p* = .013), if they are native speaker of English (*r*(14) = -.683, *p* = .007), their self-report recognition of multitasker (*r*(14) = -.683, *p* = .007), and if the participants used their dominant hand to complete the task (*r*(14) = -.631, *p* = .016). See table 16 (Appendix) for the result of correlation.

Additionally, the self-report surprise level and the facial expression checklist for surprise used in our study showed that the first warning tone was creating the surprise response but the second warning tone was not effective. 6 out of 14 participants report 5 or higher scores on the first warning tone and only one participant showed surprise at the second warning tone. Similarly, half of the participants showed surprise in their facial expression for the first warning and only 2 showed surprise at the second. Table 17 presents the descriptive summary for the responses of these two measurements.

Table 17
Descriptive Summary for the Responses of Facial Expression Checklist for Surprise and Self-Report Surprise Scale

| | Facial expression check list | | Self-report surprise scale | |
|---|---|---|---|---|
| | No | Yes | < 5 | ≥ 5 |
| First warning tone | 7 | 7 | 8 | 6 |
| Second warning tone | 12 | 2 | 13 | 1 |

*Note.* Six characteristics were used to analyze the facial expression which were the movement of their eyebrows, eyes, jaw, if there are any sudden movements, if there are any sudden noises, if participants look surprised and if they gaze in the direction of their partner. Self-report surprise scale ranges from 1 (Not at all surprised) to 7 (As surprised as one can be).

# CHAPTER 4

# DISCUSSION

In this study, we applied a physical coordination task to explore the effect of attribution bias including self-serving bias and correspondence bias on people's trust and their attribution of blame. However, the hypotheses regarding both biases were not supported. We found little evidence of the presence of the effects. Besides the investigation of these effects, we gleaned three major findings from testing the hypotheses.

First, the hypothesis regarding the engagement of the self-serving bias was not supported. However, on average, our results suggest that when people coordinate with an unfamiliar human partner and they end up in a negative result, people will be more likely to shift blame toward themselves or an unexpected environmental event compared to their human partner. This result is opposite to our hypothesis about the engagement of self-serving bias. This might be the result of the fact that self-serving bias in causal attributions appear to be weakened when people perform in groups (Zaccaro, Peterson & Walker, 1987) Also, this may be due to the similarity of the participants and the friendliness of the participants. Our sample mainly consisted of college students who are studying in Arizona State University, generally majoring in the same program and might have future interaction with each other. Besides, participants' friendliness can also influence the attribution of blame. Although participants were separated when they answered the trust and attribution of blame questionnaires, and were informed the result would not be disclosed to anyone other than the researchers involved in this experiment, the possible future connection and their friendless can still influence their answers. This finding is consistent with and can be extended to support the conclusion made by Groom, Chen, Johnson, Kara and Nass (2010). That is, friendliness affects how people attribute blame, not only when they interact with another human, but also with their machine partner. In addition, this finding may be evidence that the individual's differences and needs should be taken into account when designers are designing the machines.

Second, the results demonstrated a significant difference between post-task and initial trust in the partners when an unexpected event was involved in the physical coordination. Surprisingly, instead of experiencing a decrease of trust, people are likely to gain more trust in their partner after the unexpected event. In addition, it is interesting to note that people who have been through the coordination without the involvement of the event did not show the same pattern. This may be caused by the increased understanding of the situation they are facing including the environmental situation and the partner's behavior, and thus, possessing more trust in their partner.

Finally, the result revealed that how participants assign blame for the negative outcome is a significant predictor for how they will rate post-task trust in their partner. An implication is that which entities people assign blame to may affect the amount of trust they have in their partner. For instance, when a joint task resulted in a negative outcome, people may tend to blame themselves or their partner, moreover, this tendency can further influence the content of both ongoing and future interactions with the partner. However, the attribution of blame was only a significant predictor of post-task trust, not initial trust. This could be the result of the lack of understanding of their partner at the first impression. This conclusion is consistent with the findings in human-machine interaction and human-robot interaction that the transparency of machines and robots impacts people's trust in them (Lyons& Havig, 2014, June; Wortham & Theodorou, 2017). Furthermore, this implication can complement the previous study in the human-machine interaction about blame. Kim and Hinds (2006) argued that transparency of the machine may impact the way people assign blame to the machine. This study further suggested that transparency of the machine may not only affect people's assignment of blame but also their trust in the machines.

From the results of checking the potential confound for the relationship between trust and attribution of blame, we found that in the surprise condition: the taller the participants are, the more trust they give to their partner for both the initial trust or the post-task trust. This phenomenon may be due to the relation between the force participants use and their self-confidence to complete the task alone successfully, which may be a confound of our study. Besides the finding about the usage of force, there are multiple correlations that were found by the analysis. Although the result indicated strong significant differences, the significance of these correlations was due to the similar pattern of answer. To measure participants' demographic information, the present study used the nominal scale for gender (male, female), if participants self-reported they are native speakers of English, if they are multitaskers, and if they completed the task with their dominant hand (Yes/ No). The ratio scale was used for weight (120-160; 160-200; 200-240). The usage of these scales may lead to incorrect statistical results. Future research may use different types of questionnaires to get more precise data.

In addition, the result indicated a significant difference between the self-blame score and the warning tone-blame score in the baseline condition. This difference is foreseeable since we did not include the warning tone in the baseline condition.

Observations from the interactions among humans give insight into how we interact with robots. In a series of studies based on a research paradigm called *computers are social actors* (CASA), Nass and his colleagues have demonstrated that social rules guiding human–human interaction may apply equally to human–computer interaction, with users responding to machines as independent entities rather than as a manifestation of their human creators (Sundar & Nass 2000). Similarly, evidence demonstrated that people treat robots as social actors and robots are not always perceived by their users as technologies (Friedman, Kahn & Hagman, 2003; Lee, Park, & Song, 2005; Young, Hawkins, Sharlin & Igarashi, 2009). These studies indicate that social psychological theory can enlighten our understanding of how people interact with technologies.

Likewise, our study builds upon human to human interaction and aims to further explore the differences and similarities when people coordinate with machines. Although our hypotheses were inconclusive in the present study, we believe our work provides some evidence for the necessity for looking into the effects of attribution bias in further human-machine coordination.

## Limitation

We present several limitations for the present research. First, we used a sample of college students, which is common in past research. However, the homogeneous background of the students may be a potential confound of this study. In addition, the sample consists of students who may not be able to represent the broader population, because they are generally well-educated and exposed to technology. Therefore, we suggest future researchers collect data from a larger sample to enhance the diversity of participants' background. The second limitation was the motivation of the participants. Motivation levels may differ from individuals. We attempted to increase and standardize motivation by offering a $30 prize to the best participants with the best performance. However, we cannot confirm the incentive of our manipulation. The third limitation was that the value of the object and the risk to carry objects would likely greatly affect trust and the behavior of the participants. In this study, a box containing three bricks with 5.75 lbs and a cup with 200ml of water was used as the object to transport. However, real world applications would likely involve much more uncertain situations and objects that are much heavier, and that could possibly injure the person. Such uncertainties and risks would likely be taken into account in completing the task, which is not considered in this study. It is worth exploring if these considerations could impact human trust in their partner and their attribution of blame. Finally, this study was a part of a larger on-going study, causing other variables to affect the results received. For example, two other questionnaires were used in this study. Each of the questionnaire contains more than twenty questions and relates to other constructs of trust such as trust in automation (Jian, 2000) and interpersonal trust scale (Rotter, 1967). This may enhance the participants' mental fatigue to answer the questions and can influence participants' answers by making them answer with less self-reflection. Besides the irrelevant questionnaire, motion capture devices, which were used for a different study, might lead participants to be more aware of their movements compared to the case without the existence of them.

We anticipate the continued works in this area will improve the experimental setting by considering and adjusting for these limitations.

## Conclusion

In this study, we aimed to gain understandings on the effect of attribution biases such as self-serving bias and correspondence bias on the development of human trust in their partner in the coordination context. Based on the findings from the literature review, we predicted that in the coordination context, people would blame their partner more when the result of the coordination was negative, at the same time, their trust in the partner would decrease due to the negative outcome. Second, we assumed that if an environmental instability such as if an unexpected event occurred during the coordination which then results in a negative outcome, the levels of blame to their partner would be higher and their trust in their partner would be lower compared to the case without the environmental instability.

We designed a “microworld” transportation coordination task and conducted a between-subject experimental design study to examine the hypotheses by randomly assigning participants into either the coordination involving a designated warning tone which serves as an unexpected event or without the warning tone (which does not hinder the task).

Although the study was well-designed, our results suggested that our predictions were not supported by the collected data. The effects of the attribution biases were not observable in the present study. However, there are three findings we gleaned from the data analysis. First, we found that the friendliness of people may be a factor that affects their assignment of blame. Second, the gained understanding of the situation including the task environment and the partner’s behaviors may be a catalyst for gaining human trust in their partner. Third, the result indicated that how people assign blame for the negative outcome may influence the amount of trust they give to their partner. In addition, participants’ height might be a potential confound for research in physical coordination.

In spite of the present study being built upon physical human to human coordination, we believe these findings can enlighten our understanding of wider areas such as automation-aid scenario and coordination in general. Given the findings of this study, future investigation on the effects of attribution biases in human-machine coordination is necessary.

# REFERENCES

Agravante, D. J., Cherubini, A., Bussy, A., & Kheddar, A. (2013, November). Human-humanoid joint haptic table carrying task with height stabilization using vision. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*(pp. 4609-4614). IEEE.

Bengler, K., Zimmermann, M., Bortot, D., Kienle, M., & Damböck, D. (2012). Interaction principles for cooperative human-machine systems. it-Information Technology Methoden und innovative Anwendungen der Informatik und Informationstechnik, 54(4), 157-164.

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.

Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The effectiveness of airline pilot training for abnormal events. Human factors, 55(3), 477-485.

Cassell, J., & Bickmore, T. (2000). External manifestations of trustworthiness in the interface. *Communications of the ACM*, *43*(12), 50-56.

Chen, J. Y., & Barnes, M. J. (2012). Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, *54*(2), 157-174.

Chen, J. Y., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, *44*(1), 13-29.

Cherubini, A., Passama, R., Crosnier, A., Lasnier, A., & Fraisse, P. (2016). Collaborative manufacturing with physical human–robot interaction. *Robotics and Computer-Integrated Manufacturing*, *40*, 1-13.

Davis, W. L., & Davis, D. E. (1972). Internal-external control and attribution of responsibility for success and failure 1. *Journal of Personality*, *40*(1), 123-136.

De Michele, P. E., Gansneder, B., & Solomon, G. B. (1998). Success and failure attributions of wrestlers: Further evidence of the self-serving bias. *Journal of Sport Behavior*, 21(3), 242.

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.

de Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. Journal of Cognitive Engineering and Decision Making, 5(2), 209-231.

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. International Journal of Human-Computer Studies, 58(6), 719-735.

Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., & Yanco, H. (2013, March). Impact of robot failures and feedback on real-time trust. In Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction (pp. 251-258). IEEE Press.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. International journal of human-computer studies, 58(6), 697-718.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. Human Factors, 44(1), 79-94.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting misuse and disuse of combat identification systems. *Military Psychology*, *13*(3), 147.

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, USA.

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.

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007, May). Measurement of trust in human-robot collaboration. In *Collaborative Technologies and Systems, 2007. CTS 2007. International Symposium on* (pp. 106-114). IEEE.

Friedman, B. (1995, May). "It's the computer's fault": reasoning about computers as moral agents. In *Conference companion on Human factors in computing systems* (pp. 226-227). ACM.

Friedman, B., Kahn Jr, P. H., & Hagman, J. (2003, April). Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 273-280). ACM.

Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. Journal of Management, 38(4), 1167-1230.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. Psychological bulletin, 101(1), 75.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. Psychological bulletin, 117(1), 21.

Granados, D. F. P., Yamamoto, B. A., Kamide, H., Kinugawa, J., & Kosuge, K. (2017). Dance Teaching by a Robot: Combining Cognitive and Physical Human–Robot Interaction for Supporting the Skill Learning Process. *IEEE Robotics and Automation Letters*, *2*(3), 1452-1459.

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, *8*(3), 483-500.

Groom, V., Chen, J., Johnson, T., Kara, F. A., & Nass, C. (2010, March). Critic, compatriot, or chump?: Responses to robot blame attribution. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction* (pp. 211-218). IEEE Press.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407-434.

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cognition & Emotion*, *20*(2), 242-273.

Hung, M. C., Chang, I. C., & Hwang, H. G. (2011). Exploring academic teachers' continuance toward the web-based learning system: The role of causal attributions. *Computers & Education*, *57*(2), 1530-1543.

Itoh, M., Abe, G., & Tanaka, K. (1999). Trust in and use of automation: their dependence on occurrence patterns of malfunctions. In *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on* (Vol. 3, pp. 715-720). IEEE.

Jarrassé, N., Charalambous, T., & Burdet, E. (2012). A framework to describe, analyze and generate interactive motor behaviors. *PloS one*, *7*(11), e49945.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics, 4(1), 53-71.

Jones, E. E. (1979). The rocky road from acts to dispositions. *American psychologist*, *34*(2), 107.

Jones, E. E., & Nisbett, R. E. (1987). The actor and the observer: Divergent perceptions of the causes of behavior. In *Preparation of this paper grew out of a workshop on attribution theory held at University of California, Los Angeles, Aug 1969.* Lawrence Erlbaum Associates, Inc.

Kaniarasu, P., & Steinfeld, A. M. (2014, August). Effects of blame on trust in human robot interaction. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on* (pp. 850-855). IEEE.

Kim, T., & Hinds, P. (2006, September). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on* (pp. 80-85). IEEE.

Kochan, J. A., Breiter, E. G., & Jentsch, F. (2004, September). Surprise and unexpectedness in flying: Database reviews and analyses. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 3, pp. 335-339). Sage CA: Los Angeles, CA: SAGE Publications.

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.

Kucukyilmaz, A., & Demiris, Y. (2018). Learning Shared Control by Demonstration for Personalized Wheelchair Assistance. *IEEE Transactions on Haptics*.

Kylen, B. J. (1985). What business leaders do–before they are surprised. *Advances in strategic management*, *3*, 181-222.

Lau, R. R., & Russell, D. (1980). Attributions in the sports pages. *Journal of personality and social psychology*, 39(1), 29.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International journal of human-computer studies*, *40*(1), 153-184.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243-1270.

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.

Lerch, F. J., Prietula, M. J., & Kulik, C. T. (1997, May). The Turing effect: The nature of trust in expert systems advice. In *Expertise in context* (pp. 417-448). MIT Press.

Li, Y., Tee, K. P., Yan, R., Chan, W. L., & Wu, Y. (2016). A framework of human–robot coordination based on game theory and policy iteration. *IEEE Transactions on Robotics*, *32*(6), 1408-1418.

Lyons, J. B., & Havig, P. R. (2014, June). Transparency in a human-machine context: Approaches for fostering shared awareness/intent. In International Conference on Virtual, Augmented and Mixed Reality (pp. 181-190). Springer, Cham.

Madhavan, P., & Wiegmann, D. A. (2004, September). A new look at the dynamics of human-automation trust: Is trust in humans comparable to trust in machines?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 3, pp. 581-585). Sage CA: Los Angeles, CA: SAGE Publications.

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277-301.

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.

Merritt, S. M. (2011). Affective processes in human–automation interactions. *Human Factors*, *53*(4), 356-370.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, *50*(2), 194-210.

Meyer, W. U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. Cognition & Emotion, 5(4), 295-311.

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.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction?. *Psychological bulletin*, *82*(2), 213.

Moon, Y. (2003). Don’t blame the computer: When self-disclosure moderates the self-serving bias. *Journal of Consumer Psychology*, *13*(1-2), 125-137.

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.

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.

Morgan, T. (1992). Competence and responsibility in intelligent systems. *Artificial intelligence review*, *6*(2), 217-226.

Mörtl, A., Lawitzky, M., Kucukyilmaz, A., Sezgin, M., Basdogan, C., & Hirche, S. (2012). The role of roles: Physical cooperation between humans and robots. The International Journal of Robotics Research, 31(13), 1656-1674.

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.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429-460.

Nass, C., Moon, Y., & Carney, P. (1999). Are People Polite to Computers? Responses to Computer-Based Interviewing Systems 1. *Journal of Applied Social Psychology*, *29*(5), 1093-1109.

Parasuraman, R., & Miller, C. A. (2004). Trust and etiquette in high-criticality automated systems. Communications of the ACM, 47(4), 51-55.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230-253.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced'complacency'. *The International Journal of Aviation Psychology*, *3*(1), 1-23.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140-160.

Park, E., Jenkins, Q., & Jiang, X. (2008). Measuring trust of human operators in new generation rescue robots. In Proceedings of the JFPS International Symposium on Fluid power (Vol. 2008, No. 7-2, pp. 489-492). The Japan Fluid Power System Society.

Park, E., Kim, K. J., & Del Pobil, A. P. (2011, November). The effects of a robot instructor's positive vs. negative feedbacks on attraction and acceptance towards the robot in classroom. In *International Conference on Social Robotics* (pp. 135-141). Springer, Berlin, Heidelberg.

Pinto, J. K., Slevin, D. P., & English, B. (2009). Trust in projects: An empirical assessment of owner/contractor relationships. *International Journal of Project Management*, *27*(6), 638-648.

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.

Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014, September). Startle and surprise on the flight deck: Similarities, differences, and prevalence. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 58, No. 1, pp. 1047-1051). Sage CA: Los Angeles, CA: SAGE Publications.

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016, March). Overtrust of robots in emergency evacuation scenarios. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 101-108). IEEE Press.

Ross, L. (1977). The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process1. In *Advances in experimental social psychology* (Vol. 10, pp. 173-220). Academic Press.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust 1. Journal of personality, 35(4), 651-665.

Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, *49*(1), 76-87.

Sampson Jr, J. P. (1986). Computer technology and counseling psychology: Regression toward the machine?. The Counseling Psychologist, 14(4), 567-583.

Schraft, R. D., Meyer, C., Parlitz, C., & Helms, E. (2005, April). PowerMate–A safe and intuitive robot assistant for handling and assembly tasks. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on* (pp. 4074-4079). IEEE.

Schützwohl, A., & Borgstedt, K. (2005). The processing of affectively valenced stimuli: The role of surprise. Cognition & Emotion, 19(4), 583-600.

Sharek, Z., Swift, S., Gino, F., & Moore, D. (2010). Not As Big As It Looks: Attribution Errors in the Perceptual Domain. *ACR North American Advances*.

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor. *Communication research*, *27*(6), 683-703.

Swann Jr, W. B., & Schroeder, D. G. (1995). The search for beauty and truth: A framework for understanding reactions to evaluations. *Personality and Social Psychology Bulletin*, *21*(12), 1307-1318.

Unhelkar, V. V., Siu, H. C., & Shah, J. A. (2014, March). Comparative performance of human and mobile robotic assistants in collaborative fetch-and-deliver tasks. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 82-89). ACM.

van der Woerdt, S., & Haselager, P. (2017). When robots appear to have a mind: the human perception of machine agency and responsibility. *New Ideas in Psychology*.

Verberne, F. M., Ham, J., & Midden, C. J. (2012). Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human factors*, *54*(5), 799-810.

Vilaza, G. N., Haselager, W. F. F., Campos, A. M. C., & Vuurpijl, L. (2014). Using games to investigate sense of agency and attribution of responsibility. *Proceedings of the 2014 SBGames (SBgames 2014), SBC, Porte Alegre*.

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.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. Human factors, 51(3), 281-291.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. guilford Press.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). Engineering psychology & human performance. Psychology Press.

Wisse, F. Effects of Adaptive Support on Team Performance by advising Human Reliance Decision Making and Adaptive Automation.

Woerdt, S., & Haselager, W. F. G. (2016). Lack of effort or lack of ability? Robot failures and human perception of agency and responsibility.

Wortham, R. H., & Theodorou, A. (2017). Robot transparency, trust and utility. Connection Science, 29(3), 242-248.

You, S., Nie, J., Suh, K., & Sundar, S. S. (2011, March). When the robot criticizes you...: self-serving bias in human-robot interaction. In Proceedings of the 6th international conference on Human-robot interaction (pp. 295-296). ACM.

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.

Zaccaro, S. J., Peterson, C., & Walker, S. (1987). Self-serving attributions for individual and group performance. *Social Psychology Quarterly*, 257-263.

# APPENDIX A

## TABLE TWO

Table 2

*Descriptive Statistics of Muir's Trust Questionnaires Responses for Each Condition*

| Scale | Baseline<br>(*n* = 12) | | | | Surprise<br>(*n* = 14) | | | |
|---|---|---|---|---|---|---|---|---|
| | Initial | | Post | | Initial | | Post | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Predictability | 3.42 | 1.00 | 4.25 | 0.75 | 3.71 | 1.07 | 4.43 | 1.28 |
| Dependability | 4.00 | 0.74 | 4.25 | 0.97 | 4.36 | 0.93 | 4.79 | 0.58 |
| Responsibility | 3.83 | 0.72 | 4.42 | 0.67 | 4.14 | 1.29 | 4.79 | 0.58 |
| Competence | 4.00 | 0.74 | 4.08 | 1.08 | 4.57 | 0.76 | 4.79 | 0.58 |
| Overall Trust | 4.25 | 0.75 | 4.58 | 0.67 | 4.29 | 1.14 | 4.79 | 0.58 |

*Note.* Scales of Muir's trust questionnaire range from 1 (strongly disagree) to 5 (strongly agree)

# APPENDIX B

## TABLE FOUR

Table 4

*Descriptive Statistics of Attribution of Blame Questionnaire Responses for Each Condition*

| Scales | Baseline | | | Surprise | | |
|---|---|---|---|---|---|---|
| | *N* | *M* | *SD* | *N* | *M* | *SD* |
| **Attribution of blame to self** | | | | | | |
| - I was responsible for the unsuccessful result | 10 | 3.50 | 1.78 | 12 | 3.17 | 1.64 |
| - I was to blame for the unsuccessful result | 10 | 3.00 | 2.21 | 12 | 3.33 | 1.67 |
| **Attribution of blame to the partner** | | | | | | |
| - My partner was to blame for the unsuccessful result | 9 | 2.67 | 1.50 | 12 | 3.00 | 1.76 |
| - My partner was responsible for the unsuccessful result | 10 | 2.10 | 1.45 | 12 | 2.67 | 1.83 |
| **Attribution of blame to the warning tone** | | | | | | |
| - The warning tone was to blame for the unsuccessful result | 10 | 1.10 | 0.31 | 12 | 3.17 | 2.08 |
| - The warning tone was responsible for the unsuccessful result | 10 | 1.00 | 0 | 12 | 3.17 | 1.99 |

*Note.* self-blame, partner-blame and warning tone range from 1 (strongly disagree) to 7 (strongly agree)

# APPENDIX C

## TABLE NINE

Table 9

*The Average Scores of Muir's Trust Questionnaires both Initial and Post-Task in Different Conditions*

| Baseline | | | | | | | Surprise | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Initial<br>($n$ = 12) | | | Post-task<br>($n$ = 12) | | | | Initial<br>($n$ = 14) | | | Post-task<br>($n$ = 14) | | | |
| *M* | *SD* | *SE* | *M* | *SD* | *SE* | *t* | *M* | *SD* | *SE* | *M* | *SD* | *SE* | *t* |
| 3.90 | 0.62 | 0.18 | 4.31 | 0.71 | 0.21 | -1.91 | 4.17 | 0.75 | 0.20 | 4.61 | 0.68 | 0.18 | -2.33* |

*Note.* *$p$ < 0.05

# APPENDIX D

## TABLE FIFTEEN

Table 15

The result of Spearman's Correlations Between the Initial Trust, Post-Task Trust, Attribution of Blame and the Demographic Measures for the Baseline Condition

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Initial trust | – | | | | | | | | |
| 2. Post-Task trust | .368 | – | | | | | | | |
| 3. Attribution of blame | -.558 | -.467 | – | | | | | | |
| 4. Age | .220 | -.137 | . | – | | | | | |
| 5. Height | -.101 | -.518 | .281 | .044 | – | | | | |
| 6. Native speaker of English | -.112 | .262 | -.058 | -.522 | .000 | – | | | |
| 7. Multitasker | -.220 | .137 | . | -1.00** | -.044 | .522 | – | | |
| 8. Dominant hand usage | . | . | . | . | . | . | . | – | |
| 9. Weight | -.112 | .262 | -.058 | -.522 | .000 | 1.00** | .522 | . | – |

*Note.* ** Correlation is significant at the 0.01 level (2-tailed).

# APPENDIX E

## TABLE SIXTEEN

Table 16
The Result of Spearman’s Correlations Between the Initial Trust, Post-Task Trust, Attribution of Blame and the Demographic Measures for the Surprise Condition

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Initial trust | – | | | | | | | | |
| 2. Post-Task trust | .528 | – | | | | | | | |
| 3. Attribution of blame | .296 | .211 | – | | | | | | |
| 4. Age | -.018 | -.115 | .180 | – | | | | | |
| 5. Height | .619* | .659* | .074 | -.216 | – | | | | |
| 6. Native speaker of English | .179 | .054 | . | -.471 | .153 | – | | | |
| 7. Multitasker | .179 | .054 | . | -.471 | .153 | 1.00** | – | | |
| 8. Dominant hand usage | .283 | .301 | .219 | -.603* | .390 | .782** | .782** | – | |
| 9. Weight | .182 | .164 | .462 | .644* | -.134 | -.683** | -.683** | -.631* | – |

*Note.* * Correlation is significant at the 0.05 level (2-tailed).

# APPENDIX F

# CONSENT FORM

**ARIZONA STATE UNIVERSITY**

Research Participant Information and Consent Form

Title of research study: Dynamic Modeling of Joint Object Transport
Principle Investigator: Erin Chiou, Erin.Chiou@asu.edu

Primary student researcher: Yiwei Wang, Yiwei.Wang.3@asu.edu

***Description of The Research***

You are invited to participate in a research study on joint physical coordination because you are between the age of 18-55, not pregnant or susceptible to heart disease, able to consent to participate in this study, have no previous direct interaction with the partner you are participating with, and can comfortably transport 10 lbs without aid.

The purpose of the research is to examine measures of dyadic human physical coordination in a joint object transport task. The information gathered will be used to answer research questions regarding fundamental cognitive mechanisms that govern performance of people interacting with each other. This understanding will support the design of human-robot teams to improve their joint effectiveness.

This study will include staff, students, or affiliates of Arizona State University and will take place at in the Technology Center building at the Polytechnic campus.

***What will my participation involve?***

If you decide to participate in this research you will be asked to complete a series of tasks with, or while interacting with, a partner. You will also be asked to answer a series of questions related to your perceptions of the interaction, understanding of the task, your experience with joint object transport, and some demographics information that is pertinent to evaluating the task. Your total participation will be approximately 1 hour.

***How many people will be studied?***

We expect about 60 people will participate in this research study.

***What if I consent to participate, but I change my mind later?***

If at any point you feel uncomfortable or simply changed your mind about participating, you may stop the study and leave at any time. This will not be held against you.

***Are there any risks to me?***

We do not anticipate any risks to you from participation in this study.

***Are there any benefits to me?***

We do not expect any direct benefits to you from this study. However, we hope that in the future, society will benefit from this study as a result of improved technologies including robot helpers.

***Will I be compensated for my participation?***

We will not be giving any compensation for this study. Although, if your apart of a course that offers class credit, please follow the steps below.

Prior arrangement has been made with certain class instructors, for students to receive class credit for participating in this study. If this applies to you, please initial next to the statement below and write in your course number and name.

_____ Yes, I would like to receive class credit (list approved course): ____________________________

If you do not have regular access to parking near the study site at the time of scheduling, you will be compensated for one-hour parking at the hourly lots on the Polytechnic campus. You must show your parking stub.

*If you need to withdraw prior to the end of the study, you will not receive participation credit.*

***How will my confidentiality be protected?***
Your name and contact information will be collected for the sole purpose of receiving course credit. If you choose not to receive course credit, you do not need to provide this information.

Because demographics information may be key factors in the task, we ask that you do your best to answer all questions accurately. Study responses will be kept confidential.

While there will likely be publications as a result of this study, your name and any identifiable information will not be used in any other way. Only group characteristics will be published.

***Whom should I contact if I have questions?***
You may ask questions about the research at any time. If you have questions after you leave today, you should contact the Principal Investigator Erin Chiou at Erin.Chiou@asu.edu. You may also contact the primary student researcher, Yiwei Wang at Yiwei.Wang.3@asu.edu.

The study and consent form were reviewed and approved by the Social Behavioral IRB. You may contact them at (480) 965-6788 or at research.integrity@asu.edu if:

- If your questions or concerns are not being answered by the research team.
- If you cannot reach the research team.
- If you want to talk to someone besides the research team.
- If you have questions about your rights as a research participant.
- If you want to get information or provide input about this research.

Your participation is completely voluntary. Please sign and date below if you would like to continue with the study. Thank you!

Signature: ___________________________________ Date: ___________________

# APPENDIX G

# DEBRIEF DOCUMENT

Now that you are finished, I'd like to tell you a little bit more about the study. You were told that the purpose of this study was to investigate teams' cooperation pattern in the context of competition. Actuality, we were interested in developing an algorithm for robots to collaborate with humans as well as people's trust formation during the cooperation. Also, in order to standardize the experience and motivation that every participant has in this study, all participants would be told that the group with the best performance will receive the monetary prize. However, all of the participants will be told they are failed to win the prize at the end of the study.

we apologize for not telling you the full purpose of the study at the beginning. To protect the integrity of this research, we could not fully divulge our hypotheses at the start of the experiment. I hope you can see that if participants knew exactly what we were interested in studying, they might change their answers a little bit, which would negatively affect the quality of our research conclusions.

As you know, your participation in this study is voluntary. If you so wish, you may withdraw at this point, at which time all records of your participation will be destroyed. You will not be penalized if you choose to withdraw, you'll still receive a half of course credit. Are you comfortable with us using your data? Do you have any questions? If you have questions later, you can e-mail me using the contact information provided on the consent form. If you have questions about your rights as a participant, you can e-mail the IRB using the contact info also provided on the consent form.

Finally, we ask that you don't talk about any details of the study with other students or any potential participants. If participants know the true purpose of the study ahead of time, it will skew our results, so please do not share any information about the study.

Thank you very much for your participation today. We hope you found it enjoyable. If you would like to have a copy of the results e-mailed to you, please let me know and I will take your email address.

Have a great day!

# APPENDIX H

# IRB DOCUMENT

Instructions and Notes:

- Depending on the nature of what you are doing, some sections may not be applicable to your research. If so, mark as “NA”.
- When you write a protocol, keep an electronic copy. You will need a copy if it is necessary to make changes.

**1** Protocol Title

Include the full protocol title: **Dynamic Modeling of Joint Object Transport**

**2** **Background and Objectives**

Provide the scientific or scholarly background for, rationale for, and significance of the research based on the existing literature and how will it add to existing knowledge.

- Describe the purpose of the study.
- Describe any relevant preliminary data or case studies.
- Describe any past studies that are in conjunction to this study.

Joint object transport scenarios are common in daily work. Moving furniture, assembly, and installation involve planned and emergent motor movements between two or more individuals. Studies have previously explored joint action and social connection, however there is increasing initiative to develop robots with advanced control policies that may augment performance in object transport in ways that a partner or partners might. While rich social channels facilitate synchronous movement, dynamic role allocation, and replanning, these channels in physical human-robot collaboration (PHRC) are often unsophisticated or misleading, resulting in choppy interaction and unintended consequences. The context of these situations are often dynamic and complex, involving continuously evolving environments, task demands, and physiological as well as mental states and events. As a result, infrequent or rare events that occur may represent dramatically different constraints than routine operation entails. An important emotional response from human actors in these scenarios is surprise.

Surprise is a critical factor in joint object transport in several ways. Expectations about one’s own abilities, their partner, the environment, goal and interactions with the object may all be violated by emergent signals. Some signals such as a coworker injury may shift the demands completely from transporting the physical load to another task, while the presence of a supervisor may represent acute pressures to exert more in the task or to be more careful. As these events are perceived, individuals may update their model of the context to more accurately reflect the current constraints of the system. This new model may be indirectly observable through changes in physical or physiological measures such as interaction force, motion, or through self-report.

Establishing consistent patterns in relevant measures of joint object transport following surprise may elucidate the process of recovery, avoidance, and increasing interaction stability with human dyads. This insight may aid in the development of augmentative technologies, such as assistive robots or wearable sensors that can facilitate adaptation to unexpected events that elicit surprise, as well as fulfill primary intended purposes such as increasing efficiency in routine situations.

**3** **Data Use**

Describe how the data will be used.

Examples include:

- Dissertation, Thesis, Undergraduate honors project
- Publication/journal article, conferences/presentations
- Results released to agency or organization
- Results released to participants/parents
- Results released to employer or school
- Other (describe)

The data will be used for publication in journal articles and in conference submissions and presentations. De-identified or aggregate data may also be used as part of class practicums or K-12 outreach activities that the researchers are involved in.

### 4 Inclusion and Exclusion Criteria

Describe the criteria that define who will be included or excluded in your final study sample. If you are conducting data analysis only describe what is included in the dataset you propose to use. Indicate specifically whether you will target or exclude each of the following special populations:

- Minors (individuals who are under the age of 18)
- Adults who are unable to consent
- Pregnant women
- Prisoners
- Native Americans
- Undocumented individuals

Our recruitment criteria will target adults age 18 or older, who are able to consent, and can comfortably transport items weighing 10 lbs with one arm and without assistance from technology (self-reported). We will exclude minors, adults who are unable to consent, pregnant woman, prisoners, and undocumented individuals. We will also exclude non-English speakers due to limitations of our study team and also because it is not critical for our research question, as well as exclude adults over the age of 55 due to their potential physical ability that would make this group unlikely to be part of a population that would be routinely transporting heavy objects alongside robots in the next decade.

### 5 Number of Participants

Indicate the total number of participants to be recruited and enrolled: 200 participants are estimated for this study.

### 6 Recruitment Methods

- Describe who will be doing the recruitment of participants.
- Describe when, where, and how potential participants will be identified and recruited.
- Describe and attach materials that will be used to recruit participants (attach documents or recruitment script with the application).

Only the researchers listed on the IRB protocol (trained ASU faculty and students) will be involved in the recruitment of participants. A combination of online and physical flyers listing a short description of the study and recruitment criteria will be posted in publically accessible locations, such as campus bulletin boards, Craig's list, and community centers. The researchers may also recruit through their personal networks, email lists, and other avenues in which they are familiar with the norms of solicitation in the group, or have received advanced permission from managers of those groups, such as established study subject pools. Prospective participants will be asked to contact the researchers and confirm that they meet the study criteria before scheduling will take place.

**7 Procedures Involved**

Describe all research procedures being performed, who will facilitate the procedures, and when they will be performed. Describe procedures including:

- The duration of time participants will spend in each research activity.
- The period or span of time for the collection of data, and any long term follow up.
- Surveys or questionnaires that will be administered (Attach all surveys, interview questions, scripts, data collection forms, and instructions for participants to the online application).
- Interventions and sessions (Attach supplemental materials to the online application).
- Lab procedures and tests and related instructions to participants.
- Video or audio recordings of participants.
- Previously collected data sets that that will be analyzed and identify the data source (Attach data use agreement(s) to the online application).

The researchers will be running a between-subjects design with four randomized groups. These groups will involve interaction with a human dyad or human robot dyad. Prior to participants arriving, they will fill out individual consent forms. Following consent, the study will begin by sending an email to participants guiding them to take an Interpersonal Trust survey. After that is completed and the participants arrive in the lab they will begin by being exposed to a series of simple physical coordination tasks with a partner or robot. After the training slide, they will be ked another Interpersonal Trust Survey. These tasks will be essentially jointly lifting and holding an object (10 lbs or lighter) and coordinating their actions with their partner or robot to reach an intended location. Motion-capture data that uses video and markers attached to participants' arms and the interaction force of the object will be recorded as participants complete the task. Depending on the study group:

1. the task may also involve an infrequent, low-intensity auditory tone, signalling that the interaction should be ceased until the tone ends.
2. Participants may observe an example of the infrequent, low-intensity tone.
3. Participants may have facial expression data observed and documented.

**8 Compensation or Credit**

- **Describe the amount and timing of any compensation or credit to participants.**
- **Identify the source of the funds to compensate participants**
- **Justify that the amount given to participants is reasonable.**
- **If participants are receiving course credit for participating in research, alternative assignments need to be put in place to avoid coercion.**

Students in introductory psychology or equivalent courses will receive course credit in lieu of an alternate assignment. The alternate assignment will be to write a two-page paper on dyadichuman and robot coordination, which the faculty researchers believe are commensurate with the experience and time spent completing the study.

**9 Risk to Participants**

List the reasonably foreseeable risks, discomforts, or inconveniences related to participation in the research. Consider physical, psychological, social, legal, and economic risks.

The reasonably foreseeable risks, discomforts, or unexpected inconveniences related to participation in the research are minimal. These may include: disrupting their daily routine to travel to the study site, have difficulty finding parking, potentially receiving a parking ticket if they do not comply with parking signs or our instructions, experiencing boredom with the task if they had high expectations about the dyadic coordination task, experience anxiety when asked to complete multiple tasks simultaneously, experiencing anxiety working with a partner or robot and experiencing anxiety when we collect information about their age, height or weight, and experiencing anxiety from being deceived about the potential prize. We plan to minimize these risks and discomforts by being as flexible as reasonably possible in scheduling, or in some cases rescheduling study participation, providing a map and information about parking well in advance of their scheduled time so they can prepare to travel to the study site, fully explain the task, and to communicate clearly that they may stop the study at any time, and that their name or contact information will not be connected to the data file we collect from them, and that study findings will only be reported in aggregate. We will also debrief the participant about the use of deception in this study.

## 10 Potential Benefits to Participants

Realistically describe the potential benefits that individual participants may experience from taking part in the research. Indicate if there is no direct benefit. Do **not** include benefits to society or others.

There are no other direct benefits to participants.

## 11 Privacy and Confidentiality

Describe the steps that will be taken to protect subjects' privacy interests. "Privacy interest" refers to a person's desire to place limits on with whom they interact or to whom they provide personal information. Click here for additional guidance on ASU Data Storage Guidelines.

Describe the following measures to ensure the confidentiality of data:

- Who will have access to the data?
- Where and how data will be stored (e.g. ASU secure server, ASU cloud storage, filing cabinets, etc.)?
- How long the data will be stored?
- Describe the steps that will be taken to secure the data during storage, use, and transmission. (e.g., training, authorization of access, password protection, encryption, physical controls, certificates of confidentiality, and separation of identifiers and data, etc.).
- If applicable, how will audio or video recordings will be managed and secured. Add the duration of time these recordings will be kept.
- If applicable, how will the consent, assent, and/or parental permission forms be secured. These forms should separate from the rest of the study data. Add the duration of time these forms will be kept.
- If applicable, describe how data will be linked or tracked (e.g. masterlist, contact list, reproducible participant ID, randomized ID, etc.).

If your study has previously collected data sets, describe who will be responsible for data security and monitoring.

Only the study personnel listed in the IRB will have access to the data. Data will be stored on ASU secure servers, conventionally accessible only through password-protected computers and password-protected accounts. Any paper data (i.e., consent forms with names and signatures) will be located in a locked file cabinet for up to 1 year following data collection, after which those materials will be shredded and recycled. Participant ID number counting upwards from 100 will be assigned to participants in the order that they are scheduled. While this assignment procedure is not completely random and thus potentially traceable, we believe the convenience for auditing the data that this approach allows outweighs the potential risk to any breach in participants' privacy.

**12 Consent Process**

Describe the process and procedures process you will use to obtain consent. Include a description of:

- Who will be responsible for consenting participants?
- Where will the consent process take place?
- How will consent be obtained?
- If participants who do not speak English will be enrolled, describe the process to ensure that the oral and/or written information provided to those participants will be in that language. Indicate the language that will be used by those obtaining consent. Translated consent forms should be submitted after the English is approved.

Only the study personnel listed in the IRB trained in human subject research ethics will be responsible for consenting participants. Consent will take place before participant s arrive to the study suite, they will receive the consent form via email, after they sign up for the study. Consent will involve going through a electric consent form making them aware of the potential benefits and reiterating that they may choose to end their participation at any time should they feel any discomfort. Participants must electronically sign the consent form and send it back to researchers prior to arrival at the study site. If participants do not fill out consent form, they will be removed from the study and will not be able to participate. All study activities will be conducted in English.

**13 Training**

Provide the date(s) the members of the research team have completed the CITI training for human participants. This training must be taken within the last 4 years. Additional information can be found at: Training.

**Erin Chiou, 23-Aug-2016**
**Wenlong Zhang, 15-Mar-2017**
**Yiwei Wang, 31-Jul-2017**
**Pouria Salehi, 18-Oct-2016**
**Glenn Lematta 27-Jan-2016**
**Chi-Ping Hsiung 1-Sep-2017**
**Kyleigh Rahm 20-Sep-2017**
**Alex Shaw 24-Apr-2017**