5-2023

Trust is Not Enough: Examining the Role of Distrust in Human-Autonomy Teams

Sydney Begerowski
sbegero@clemson.edu

TRUST IS NOT ENOUGH: EXAMINING THE ROLE OF DISTRUST IN HUMAN-AUTONOMY TEAMS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Applied Psychology

by
Sydney Begerowski
May 2023

Accepted by:
Dr. Marissa L. Shuffler, Committee Chair
Dr. Laine Mears
Dr. Patrick Rosopa
Dr. Allison Traylor

ABSTRACT

As automation solutions in manufacturing grow more accessible, there are consistent calls to augment capabilities of humans through the use of autonomous agents, leading to human-autonomy teams (HATs). Many constructs from the human-human teaming literatures are being studied in the context of HATs, such as affective emergent states. Among these, trust has been demonstrated to play a critical role in both human teams and HATs, particularly when considering the reliability of the agent performance. However, the HAT literature fails to account for the distinction between trust and distrust. Consequently, this study investigates the effects of both trust and distrust in HATs in order to broaden the current understanding of trust dynamics in HATs and improve team functioning. Findings were inclusive, but a path forward was discussed regarding self-report and unobtrusive measures of trust and distrust in HATs.

TABLE OF CONTENTS

Table of Contents (Continued) Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

The progression toward Industry 4.0 - a technological revolution focused on the adoption of real time data analytics- is drastically transforming core business processes in nearly all industries (Frank et al., 2019). The manifestation of Industry 4.0 within manufacturing (i.e., Smart Manufacturing) emphasizes deeply blended cyber and physical systems that address increasingly complex demands for manufacturers (Kang et al., 2016). While manufacturing processes improve and automation solutions continuously grow cheaper, the products they aim to create also rapidly grow in complexity. Given this increase in complexity of assembly operations, automation is often impractical. Instead, manufacturers must look to human labor to meet these demands, which constitutes an enormous cost for manufacturers. Augmenting the abilities of human laborers with cognitive or physical assistive technologies (i.e., autonomous agents) allows manufacturers to see the benefit of increased productivity in addition to allowing workers to retain value and relevance in the face of technological development (Welfare et al., 2019).

Integrating autonomous agents into main line production processes as a method for augmenting human capabilities will require complex interaction, coordination, and communication between the autonomous agent and humans, leading to human-autonomy teams (HATs). However, research regarding HATs is severely lacking. Current use of physical agents is often accomplished by using caged work cells to physically isolate the agents from humans (Siebert-Evenstone et al., 2021). Consequently, collaboration between the worker and robot is limited as most interactions are sequential in order to minimize shared physical space and preserve worker safety (Gleirscher et al., 2020). Moreover, research assessing human-autonomy teams often emphasizes a dyadic interaction rather than team (i.e., one human and one agent; Demir et al., 2017). Given these limitations, there is a growing need to understand and model collaboration among humans and autonomous agents as a new Smart Manufacturing era is evolving before our eyes. Namely, these agents appear to be undergoing a transition from tool to teammate: rather than functioning as a resource to support teams, artificial agents are expected to function as members of the team. As such, there is a critical need to understand how artificial agents can and will interface with human team members, and how fundamental concepts of human-human teaming will apply to HATs.

In the human-human teaming literature, affective states such as trust, cohesion, and efficacy play a fundamental role in effective teaming (Mathieu & Rapp, 2009). Trust is one of the most frequently studied team states, with relationships to performance, interpersonal relations, and other affective emergent states (de Jong et al., 2017). Given this importance and the historical role trust has played in implementing technology, trust is one affective state that has received considerable attention in HATs (Ghazizadeh et al., 2012; Glikson & W, 2020; Schaefer et al., 2016). In fact, promoting human trust in HATs has been identified as a critical step in successfully augmenting work in manufacturing, particularly when it comes to reliability of agent performance (Robinette et al., 2016; Welfare et al., 2019). However, there is one critical distinction in the human-human teaming literature that has not yet been considered in the HAT literature: the distinction of trust and distrust as two concepts rather than a continuum. Distrust (i.e., confident negative expectations regarding another's conduct) exists independently of trust (i.e., confident positive expectations regarding another's conduct) and is related to various teaming outcomes as well (e.g., CHO, 2006; Lewicki et al., 1998; Lowry et al., 2015).

This study seeks to not only model existing relationships between reliability of agent performance and perceived trust in agents, but to also expand the notion of trust dynamics in HATs to include that of distrust. I begin by reviewing what constitutes a human-autonomy team as well as necessary factors to define an agent as a teammate. Next, I briefly review the role of affective states in teams. I then detail the delineation between trust and distrust, reviewing the literature on trust in HATs and demonstrating the gap around distrust. Finally, I propose a design to replicate the known relationship between trust and reliability of performance while addressing the gap on distrust.

CHAPTER TWO

LITERATURE REVIEW

Human-Autonomy Teams

Teams represent two or more individuals interdependently working toward a common goal (Salas et al., 1992). The inclusion of an artificial agent into the team is not enough to constitute a human-autonomy team. Synthesizing over three decades of research on human-autonomy teaming, O'Neill et al. (2020) put forward specific criteria that must be met to be considered a human-autonomy team. Failure to meet these criteria produces a reliance on the technology as a tool or resource to enhance teamwork, rather than the agent functioning as a teammate. As such, it is imperative to understand these criteria to properly apply constructs from the human teaming literature.

First, the agent must demonstrate a degree of interdependence when working with other teammates and across tasks (O'Neill et al., 2020; Walliser et al., 2017). Second, the agent must have a degree of agency that demonstrates independence of actions (O'Neill et al., 2020; Wynne & Lyons, 2018). These criteria can be better understood when assessing the level of automation (LoA) of the agent. LoA conceptualizes both physical and cognitive abilities of an agent across a continuum, ranging from fully manual to fully automatic (Fasth et al., 2011). For example, an agent low in autonomy would rely on the human to make all decisions and actions, with the human functioning as an operator and the agent unable to complete actions without direction from the human. On the other end, an agent high in autonomy can make decisions and act on these decisions without the need for human approval. Consequently, agents low in autonomy function more as a tool whereas agents high in autonomy resemble a teammate due to the capability to manage interdependencies within the team and do so in an agentic manner. When these conditions are met, humans demonstrate politeness towards agents, use notions of self and other, and even adjust to programmed personalities of the agent (Nass et al., 1995; O'Neill et al., 2020). In turn, this perception of a teammate begins to lead to more affective states that resemble key emergent states in teams, such as trust, perceived efficacy and competence, and cohesion.

Affective Emergent States in Teams

The ABCs (i.e., affect, behavior, cognition) of teamwork is a common paradigm used to explain processes and emergent states that function as critical enablers of team performance (Mathieu et al., 2019). Team behaviors refer to team processes, which involve individuals' interdependent actions that translate inputs into outputs and lead to the emergence of team states (e.g., affect, cognition; Bell et al., 2018; Marks et al., 2001). Whereas team cognition reflects how team knowledge is mentally organized, represented, and distributed within the team, affective states represent feelings, sentiments, and attitudes within teams.

There are several affective states that influence team functioning. Collective efficacy refers to a team member's assessment of the team's collective ability to perform their given tasks (Bandura, 1997). This extends the notion of self-efficacy to the team level, reflecting an individual's confidence in the larger team to achieve group goals. Cohesion refers to member attraction to the group and is typically broken into task (i.e., shared commitment to the group task) and social (i.e., shared commitment due to positive interpersonal relations) dimensions (Brawley et al., 1987; Mullen & Copper, 1994). Meta-analyses have linked both collective efficacy and cohesion to increased performance in teams (Mullen & Copper, 1994; Stajkovic et al., 2009). Perhaps the most prominent in the teaming literature, trust is a team emergent state that has been demonstrated to directly influence team functioning in addition to influencing other affective states (de Jong et al., 2016; Mathieu et al., 2017).

Delineation of Trust and Distrust

Trust continues to be one of the most frequently studied affective states in organizational psychology (de Jong et al., 2017). Trust in a teammate refers to confident positive expectations regarding another's conduct (Lewicki et al., 1998). This is represented on a continuum, with high trust being characterized by hope, faith, confidence, and assurance in others and low trust representing lack of these characteristics as well as hesitance towards others. Fostering trust in teams is critical to enhancing team effectiveness. Meta-analyses suggest trust has an above average impact on team performance (de Jong et al., 2016; Morrissette & Kisamore, 2020). Additionally, high trust leads to better cooperation in teams (Balliet & van Lange, 2013), more satisfaction within the team (Chou et al., 2008), increased knowledge sharing among team members (Szulanski et al., 2004), and many other factors critical to effective team functioning (see Costa et al., 2018 for review).

Team and organizational researchers make a critical distinction between trust and distrust. That is, the absence of trust does not reflect distrust. Rather, distrust refers to confident negative expectations regarding another's conduct (Lewicki et al., 1998).

Distrust is also represented on a continuum. High distrust is characterized by fear, skepticism, and cynicism whereas low distrust represents a lack of these as well as low monitoring of others. Combined with trust, there are four ways in which trust can manifest (depicted in Figure 1). These four profiles are critical to consider in teams; it cannot be assumed that "the positive predictors of trust would necessarily be negative predictors of distrust, or that the positive consequences of trust would necessarily be influenced negatively by increased distrust" (Lewicki et al., 1998). Distrust has been linked to having a larger effect on an individual's decision to maintain or switch relations with another teammate than a simple lack of trust (Hardin, 1993). Moreover, distrust is related to increased competition between team members as well as increased accountability within teams, suggesting a need for balance between trust and distrust in teams (Alder, 2004; Lowry et al., 2015). For example, enhancing trust is essential in promoting lasting relationships; however, lowering distrust is key to encouraging individuals to disclose personal information in teams (Cho, 2006). Thus, the distinction between trust and distrust and how each manifests in teams can better advance our understanding of mechanisms underlying team functioning.

Trust and Distrust in HATs

Given the importance of trust in teams and the role of trust in implementing technology, trust has received considerable attention in HATs. Similar relationships in human teams have been found, with trust in HATs being positively related to performance (McNeese et al., 2021). Moreover, researchers are beginning to delineate between types of trust (e.g., initial, continuous, swift) in HATs in addition to antecedents of trust (e.g., transparency, immediacy behaviors, reliability). While an in-depth review of each form is beyond the scope of this paper (see Glikson & Wooley, 2020 for a more comprehensive review), I focus on one antecedent that appears prominently in the HAT literature: reliability of agent performance.

Reliability of agent performance refers to the extent to which the agent exhibits the same and expected behavior over time (Hoff & Bashir, 2015). Prior studies consistently demonstrate that reliability of agent performance is a significant predictor of trust in HATs (Hancock et al., 2011; Schaefer et al., 2016). However, there is no investigation into the relationship between reliability of agent performance and distrust. In fact, recent meta-analyses on trust in HATs (Hancock et al., 2011; Schaefer et al., 2016) fail to merely mention the construct of distrust. Distrust was mentioned in recent systematic reviews (Glikson & Wooley, 2020) but was operationalized as a lack of trust rather than a distinct construct. The National Academies of Sciences has recently called for further investigation into the role of distrust in HATs (i.e., Research Objective 7-4), stating the area is severely under researched with a growing need to view “trust and distrust as separate, simultaneously operating concepts” (National Academies of Sciences, 2021). Understanding the role of distrust as well as potential antecedents will be critical in enhancing the effectiveness of HATs going forward. Similar to trust, reliability of performance is also likely to influence distrust, particularly within the dimensions of cynicism and monitoring. As such, there is a significant need to better understand this relationship in HATs.

Present Study

This study seeks to expand the current understanding of trust dynamics in HATs, particularly in the manufacturing context. The purpose is two-fold. First, I seek to understand how agent reliability influences team dynamics when trust and distrust are separated into distinct constructs. In terms of trust, research suggests that agent reliability should positively impact trust perceptions. Specifically, I hypothesize that:

H1: High levels of reliability of agent performance will lead to higher levels of team trust.

On the other hand, distrust is associated with negative expectations around team members' behaviors. Given the emphasis on weariness and watchfulness associated with distrust, lower reliability of the agent (e.g., more mistakes) will likely produce a need for vigilance in human teammates. As such, my prediction is that:

H2: High levels of reliability of agent performance will lead to lower levels of team distrust.

The second aim of the study is to explore how trust and distrust in HATs uniquely contribute to different performance dimensions. In manufacturing, producing product quickly and accurately is critical, leading to the use of accuracy and cycle time as some of the most common key performance indicators in assembly teams (Chen, 2008). Additionally, vigilance (i.e., the ability to maintain focus and remain alert for extended durations; Davies & Parasuraman,1982) is becoming increasingly important in HATs as human counterparts must not overly rely on agents in an effort to reduce cognitive load (Tag et al., 2023). Given the conflicting nature of these indicators (e.g., speed-accuracy tradeoff; Heitz, 2014), each performance dimension will likely have a unique relationship with trust and distrust in HATs. Teams that are high in trust and low in distrust are typically characterized by pursuing more opportunities, exhibiting high interdependence, and monitoring one another less frequently. Consequently, I predict that:

> **H3:** Teams with higher trust perceptions in the agent and lower distrust perceptions in the agent will perform fastest.

To perform accurately, the notions of monitoring behaviors and verifying progress are likely to be important. That is, team members need to be confident in each other's abilities to complete the task (e.g., high trust). While some levels of distrust prompt desirable verification behaviors, too much distrust can lead to paranoia and fear within the team, resulting in the following hypothesis:

> **H4:** Teams with higher trust perceptions in the agent and moderate distrust perceptions in the agent will have the highest accuracy.

Finally, vigilance is a critical performance indicator such that it can signify the extent to which humans may over-rely on other teammates, especially autonomous agents. Higher levels of distrust are characterized by cynicism and weariness, and when coupled with lower levels of trust, can lead to extremely managed interdependence and overt watchfulness across team members. As such, I hypothesize that:

> **H5:** Teams with lower trust perceptions in the agent and higher distrust perceptions in the agent will demonstrate the highest vigilance.

CHAPTER THREE

METHODS

Participants

Seventy-eight students from a public university in the southern United States participated in the study. The average age was 23 years (*SD* = 4.87 years, range 18 years – 38 years). Additional descriptive information of the study can be found in Table 1. Participants were compensated in the form of Amazon Gift Cards or participatory research credits through SONA Systems.

Design

Teams of two humans and one assistive agent completed an assembly task of building three full carts. One cart was divided into three sub-assemblies: upper frame, lower frame, and body. Teams were instructed to build three carts sequentially, so they were required to complete one full build before moving onto the next cart. The role of the agent was to monitor progress of the human teammates. As the human teammates began to conclude one build (i.e., one frame), the agent would then send the next kit needed for assembly. If the participants desired a part other than what was sent, they were able to request the desired part through the agent interface. The new part was then sent by the agent after the button was pushed. Reliability of agent performance was manipulated across three levels.

Procedure

Prior to arrival, participants were required to consent to the experiment and complete a pre-survey. Upon arrival, participants were briefed on the experiment task: to build three carts with the assistance of an autonomous agent. They completed a two-part training to familiar themselves with Funphix materials and agent interaction. During training, participants practiced creating a frame using four Funphix tubes, four elbow connectors, and eight screws, allowing participants to gain familiarity with the physical materials needed to build more complex designs during the experiment. Participants were also familiarized with the agent during training by practicing using the agent interface and observing the agent sending a kit down the assembly line. They were also shown the cameras that function as the visual system for the agent and were informed of the systems monitoring functionality of the cognitive agent. At the end of training, participants took a quiz that assessed the main learning objectives to ensure understanding of the experiment and materials.

Once training was complete, participants hit the “initialize” button on the interface to begin the experiment. The team would then build three carts as part of the experiment. Once the task concluded, the participants engaged in a post-survey.

Manipulations and Measures

Reliability of Performance

Each team was randomly assigned to one of three conditions: low, moderate, and high reliability. The agent would send an incorrectly colored part as a mistake, requiring teams to identify the error and request the correct color. In the low reliability condition, the agent made three mistakes. During the third build, it sent a yellow tube instead of a green tube. During the fifth build, it sent a red tube instead of a yellow tube. During the seventh build, it sent a green tube instead of a blue. In the moderate reliability condition, the agent made two mistakes: the same mistakes at build five and seven. The agent only made one mistake (build 7) during high reliability. The introduced errors are summarized in Figure 2.

Trust and Distrust

Trust and distrust were measured using a 16-item scale with a 6-point Likert response format (Wildman et al., 2009). Participants were asked each question twice, with the referent changing to focus the item on the human teammate (e.g., to what extent did you feel assured your *human* teammate made intelligent decisions?) and agent teammate (e.g., to what extent did you feel assured your *agent* teammate made intelligent decisions?). The full survey can be found in Appendix A.

In terms of measuring trust and distrust in the human counterpart, reliability was sufficient ($\alpha = .87$); however, a confirmatory factor analysis revealed poor model fit (RMSEA = 0.161, CFI = 0.811, $\chi 2 = 209.39$). These findings were consistent for the agent counterpart as well, with satisfactory reliability ($\alpha = .93$) but poor model fit (RMSEA = 0.135, CFI = 0.855, $\chi 2 = 166.59$). Consequently, the results in the following section should be interpreted with caution.

Build Time

Performance was broken into three distinct dimensions: build time, accuracy, and vigilance. Build time was calculated in minutes based on the time taken to complete three full carts, from the send the participants hit “Initialize” to when the third cart was placed in a space designed for completed carts.

Accuracy

Accuracy reflected the extent to which teams built carts correctly. A percentage of correctly colored sub-assemblies (out of nine) was calculated, as well as percentage of screws used (out of thirty-four). The two percentages were then averaged to create a composite accuracy score.

Vigilance

Vigilance reflected the ability to stay engaged in the building task throughout the experiment. This was operationalized as a pass-fail detection of the final error in the seventh kit.

Materials

Funphix

Funphix were used as materials to build the carts. Parts included tubes (long and short), connectors (three-way, four-way, T-style), wheels, panels, screws, and keys. Colors included black, blue, green, yellow, and red. The parts were separated into three kits: upper frame, lower frame, and body. Each kit contained only the relevant parts for that portion of the build. Each kit also came with a specification sheet that provided number and color requirements for each build (Appendix B).

Assistive Agent

To model assembly processes, a mock assembly line was used to send kits and parts to participants from a designated warehouse area. An assistive agent with both physical and cognitive capabilities functioned as the third team member. While a Wizard of Oz approach was used to conduct the actual experiment, participants were shown a UR-16 completing a pick and place mechanism of a kit during training to reinforce the presence of an agent. The physical abilities were programmed in advance of the experiment, but the cognitive capabilities of systems monitoring were implemented using a Wizard of Oz approach. During the experiment, the “agent” (i.e., a trained experimenter) sent the next kit needed down the assembly line as the team concluded building the kit the team presently has, performing systems monitoring and anticipatory behaviors. The “agent” had an interface with buttons for each possible part. If the team desired a different part than sent, participants could indicate which they would like to receive instead by pushing the button that corresponds to the desired kit or part. The “agent” would then send participants the exact kit or part requested.

CHAPTER FOUR

RESULTS

Table 2 provides an overview of trust, distrust, and the performance outcomes while Table 3 summarizes these variables across each manipulation.

Reliability of Performance on Trust and Distrust

Given that an increase in reliability was associated with one less error per condition, reliability of agent performance was treated as a continuous variable (high = 8 correct behaviors, moderate = 7 correct, low = 6 correct). Accordingly, I assessed the correlation between reliability and trust, and reliability and distrust to test Hypothesis 1 and 2, respectively. Reliability of agent performance and team trust were found to be slightly, positively correlated, $r(37) = .354$, $p = .03$, supporting Hypothesis 1. However, against Hypothesis 2, reliability of agent performance was not related to distrust in teams, $r(37) = .02$, $p = .91$. Figure 3 depicts team trust and distrust across each condition.

Trust and Distrust on Performance

In order to test if trust and distrust impact the time it takes for teams to build carts (H3), a multiple regression was conducted. Average team trust, average team distrust, and an interaction between trust and distrust were used to predict build time while controlling for condition. The model was not significant, $F(4, 34) = 2.04$, $p = .11$. Figure 4 depicts build time across conditions.

A multiple regression was also conducted to assess the relationship between average team trust, average team distrust, and an interaction term while controlling for condition effects. The overall model was significant, $F(4, 34) = 4.11$, $p < .01$. However, it was found that only experiment condition predicted accuracy (β = -15.19, p < .01). Team trust (β = 1.98, p = .96), team distrust (β = 22.20, p = .96), and the interaction term (β = -2.38, p = .91) were not significant. This suggests that, as agent reliability increased, the team's accuracy decreased. Against Hypothesis 4, Team trust and team distrust did not play a significant role. Figure 5 depicts accuracy across conditions.

Finally, a logistic regression using average team trust, average team distrust, and an interaction between the two while controlling for condition membership was conducted to assess for vigilance in teams. The model was statistically significant, χ^2 (4) = 17.75, p < .01, with condition having a negative effect on vigilance (β = -1.92, p < .01). However, team trust (β = .95, p = .87) did not play a significant role, nor did team distrust (β = .50, p = .97) or an interaction between the two (β = .22, p = .93), against Hypothesis 5. More specifically, the odds-ratio indicates for a 1-unit increase in reliability (e.g., one more correct behavior), the odds of success (e.g., detecting the final error) decreased by 6.84.

Exploratory Analyses

Trust Profiles

The prior analyses used team trust and distrust as continuous predictors which is consistent with uses of the scales in the literature (e.g., Griggs et al., 2020; Lazzara et al., 2015; Thayer, 2015). However, when Lewicki and colleagues (1998) introduced the dimensional framework for trust and distrust, it was presented across four trust profiles. As such, I also identified the emergence of trust profiles within the teams in order to adequately assess the roles of trust and distrust in HATs. To create team profiles, I used the k-means clustering algorithm (Anderberg, 1973) and identified an optimal number of four clusters using the elbow method (Joshi & Nalwade, 2013). The four profiles are detailed in Table 4.

These four profiles were then used as predictors of performance in place of average measures while still controlling for condition membership. The findings were consistent with prior analyses. For build time, the model was not significant, $F(4, 34) = 1.30$, $p = .29$. While the model was significant for accuracy $F(4, 34) = 3.96$, $p < .01$, only condition had a significant effect ($\beta = -14.73$, $p < .01$). Similarly, the model was also significant for vigilance, χ^2 (2) = 17.28, $p < .001$. However, only condition had a significant effect ($\beta = -1.89$, $p < .01$).

Trust Behaviors

The presented results suggest that there are significant differences in accuracy and vigilance across conditions; however, the self-report trust and distrust measures are not adequate predictors of these differences. Given limitations of self-report data and calls for more dynamic team measures (e.g., Feitosa et al., 2020), I conducted exploratory analyses into the role of behaviors modeling trust in the agent. Prior to building each kit, participants were explicitly instructed to verify the contents of each kit. Whether or not they engaged in verifying the specifications was tracked as an indicator of trust and distrust in the agent's behavior. Table 5 details the number of teams that verified contents over time and across conditions.

For accuracy, a multiple regression was conducted with the number of checking behaviors and condition as predictors and was significant, $F(2,31) = 23.03$, $p < .000$. It was found that checking behaviors significantly predicted accuracy (β = 6.56, p < .000) while condition negatively predicted behaviors (β = -12.35, p = .001). For vigilance, a logistic regression model was conducted with specifications checked prior to the seventh kit and condition as predictors. The model was significant, χ^2 (2) = 14.74, p = .00. Both the number of kits checked prior to the seventh kit (β = .81, p = .04) and condition (β = -1.71, p < .01) were significant predictors of vigilance. This suggests that, holding condition constant, the odds of successfully detecting the final error increase by a factor of 2.26 when increasing checking behaviors by one unit. On the other hand, the odds-ratio indicates for a 1-unit increase in reliability (e.g., one more correct behavior), the odds of success (e.g., detecting the final error) decreased by 5.55.

CHAPTER FIVE

DISCUSSION

In this study, I aimed to extend the current understanding of trust dynamics in human-autonomy teams by assessing the role of distrust independently of trust. Despite prior research in human-human teams demonstrating the importance of delineating trust from distrust, findings from this study are inconclusive. First, the trust and distrust were not supported as distinct constructs as defined by Wildman and colleges (2009). There are several reasons that this may have occurred. Data may be limited by sample size (N = 36), with smaller sample sizes typically resulting in oversensitive RMSEA estimates and more likelihood of negative variance estimates (e.g., Haywood cases; Kyriazos, 2018). Additionally, trust and distrust are often conceptualized across two domains: intent and competence. While these are more intuitive for human team members, it may be difficult to delineate intent and competence across both trust and distrust in autonomous teammates (e.g., what does it mean to be wary of an agent's intent vs. competence, if possible?).

Additionally, agent reliability only slightly influenced team trust while having no effect on team distrust. Moreover, trust and distrust did not significantly predict performance across the three domains (i.e., time, accuracy, vigilance). Given the sample size and poor model fit, these findings are likely due to poor construct measurement. Another limitation is likely due to the short nature of lab-based studies, which limits the evolution of team processes (Kozlowski, 2015). The extent to which a participant's trust felt violated due to the agent making a mistake likely does not reflect the gravity of a similar situation for a worker on a manufacturing line that interacts with the agent regularly. However, condition did play a significant role in predicting accuracy and vigilance. This suggests that, over time, human behavior was changing in response to the reliability of the agent. While the self-report measures of trust failed to capture these nuances, exploratory analyses revealed a promising path forward into identifying behavioral markers that may signify trust and distrust in HATs.

Implications

First and foremost, this paper demonstrates a significant need to evaluate trust and distrust as distinct constructs in HATs, echoing various calls in the literature (e.g., National Academies of Sciences, 2021). Future research should explore the extent to which the factor structure modeled from human-human teams is generalizable to HATs. Cognitive trust in autonomous agents is well-studied (Glikson & Wooley, 2020) and often encompasses competence of agent abilities. However, understand the extent to which an individual can infer an agent's intent, and then develop trust and distrust in response to this affective judgment, will likely be more challenging.

Additionally, this study demonstrates the importance of going beyond the self-report data of trust and distrust. Unobtrusive markers of trust may more accurately reflect the dynamic nature of affective states as they unfold over a team's lifespan. Namely, identification of behaviors across the four trust profiles can create a framework for studying evolution of trust and distrust within HATs. For example, the act of checking specifications served as a "trust but verify" behavior, demonstrating moderate levels of distrust balanced with higher levels of trust.

Finally, this paper demonstrates a need to understand various operationalizations of trust and distrust in the team. While aggregated approaches are common, the cluster analysis revealed various team-level profiles that offer a different perspective on trust dynamics in teams. Namely, more research is needed to understand individual profile (in)congruence. For example, if one team member follows the trust but verify mindset (e.g., profile four) while another follows an arms-length transaction approach (e.g., profile one), there is a question of whether behaviors consistent with one profile dominate over that of another. This could provide further guidance on team composition in HATs to ensure adequate representation across to maximize performance.

Finally, determining the optimal balance of trust and distrust in manufacturing HATs can help avoid overreliance on technology while optimizing performance. High levels of trust in faulty technology are known to lead to over-trust and misuse of the technology over time, posing safety risks and leading to undesirable outcomes such as lower performance (Hoff & Bashir, 2015). In fact, failure to detect seemingly minor errors on the assembly line can lead to costly recalls for the organization (Rupp, 2004). This is especially important given that autonomous agents may reach extremely high levels of reliability; however, achieving perfect reliability of agent performance is unobtainable (Foroughi et al., 2021). Findings from this study indicate that there is a time point in which teams begin to rely too heavily on the agent, failing to detect errors made later in the task. The key to avoiding over-trust may not be lowering levels of trust but rather fostering moderate levels of distrust. In doing so, team members may be more likely to trust but verify work and monitor vulnerabilities within team functioning. While this study only functions as an initial investigation into the role of distrust in HATs, it paves way for future empirical investigations to determine the optimal balance of trust and distrust needed to promote effective human-autonomy team functioning.

REFERENCES

Alder, T. R. (2004). An Exploratory Study of Team Perceptions of Trust and Distrust in a Contract Negotiation Setting. *Journal of Behavioral and Applied Management*, *5*(3).

Anderberg, M. R. (1973). The broad view of cluster analysis. *Cluster analysis for applications, 1(1),* 1-9.

Balliet, D., & van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. https://doi.org/10.1037/a0030939

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W H Freeman/Times Books/ Henry Holt & Co.

Bell, S. T., Brown, S. G., Colaneri, A., & Outland, N. (2018). Team composition and the ABCs of teamwork. *American Psychologist*, *73*(4). https://doi.org/10.1037/amp0000305

Brawley, L. R., Carron, A. v., & Widmeyer, W. N. (1987). Assessing the Cohesion of Teams: Validity of the Group Environment Questionnaire. *Journal of Sport Psychology*, *9*(3). https://doi.org/10.1123/jsp.9.3.275

Chen, C.-C. (2008). An objective-oriented and product-line-based manufacturing performance measurement. *International Journal of Production Economics*, *112*(1), 380–390. https://doi.org/10.1016/j.ijpe.2007.03.016

Cho, J. (2006). The mechanism of trust and distrust formation and their relational outcomes. *Journal of Retailing*, *82*(1), 25–35. https://doi.org/10.1016/j.jretai.2005.11.002

Chou, L.-F., Wang, A.-C., Wang, T.-Y., Huang, M.-P., & Cheng, B.-S. (2008). Shared work values and team member effectiveness: The mediation of trustfulness and trustworthiness. *Human Relations*, *61*(12), 1713–1742. https://doi.org/10.1177/0018726708098083

Costa, A. C., Fulmer, C. A., & Anderson, N. R. (2018). Trust in work teams: An integrative review, multilevel model, and future directions. *Journal of Organizational Behavior*, *39*(2), 169–184. https://doi.org/10.1002/job.2213

Davies, D. R., & Parasuraman, R. (1982). The psychology of vigilance. *Academic Press.*

de Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*, *101*(8), 1134–1150. https://doi.org/10.1037/apl0000110

de Jong, B. A., Kroon, D. P., & Schilke, O. (2017). The future of organizational trust research: A content-analytic synthesis of scholarly recommendations and review of recent developments. In P. A. M. van Lange, B. Rockenbach, & T. Yamagishi (Eds.), *Trust in social dilemmas* (pp. 173–194). Oxford University Press.

Demir, M., McNeese, N. J., & Cooke, N. J. (2017). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, *46*, 3–12. https://doi.org/10.1016/j.cogsys.2016.11.003

Fasth, A., Mattsson, S., Fassberg, T., Stahre, J., Hoog, S., Sterner, M., & Andersson, T. (2011, May). Development of production cells with regard to physical and cognitive automation: A decade of evolution. *2011 IEEE International Symposium on Assembly and Manufacturing (ISAM)*. https://doi.org/10.1109/ISAM.2011.5942316

Foroughi, C. K., Devlin, S., Pak, R., Brown, N. L., Sibley, C., & Coyne, J. T. (2021). Near-Perfect Automation: Investigating Performance, Trust, and Visual Attention Allocation.

Human Factors: The Journal of the Human Factors and Ergonomics Society, 001872082110328. https://doi.org/10.1177/00187208211032889

Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, *210*, 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work*, *14*(1), 39–49. https://doi.org/10.1007/s10111-011-0194-3

Gleirscher, M., Johnson, N., Karachristou, P., Calinescu, R., Law, J., & Clark, J. (2020). Challenges in the Safety-Security Co-Assurance of Collaborative Industrial Robots. *Cornell University*. https://doi.org/10.1007/978-3-030-78513-0

Glikson, E., & W, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2). https://doi.org/10.5465/annals.2018.0057

Griggs, A. C., Rickel, E. A., Lazzara, E. H., & Frederick Embry-Riddle, C. M. (2020, December). Face-to-face and online friendships: examining differences in trust and distrust based on modality. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 64, No. 1, pp. 1333-1337).* Sage CA: Los Angeles, CA: SAGE Publications.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A Meta-Analysis of Factors Affecting Trust in Human-Robot

Interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Hardin, R. (1993). The Street-Level Epistemology of Trust. *Politics & Society*, *21*(4), 505–529. https://doi.org/10.1177/0032329293021004006

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in neuroscience, 8,* 150. https://doi.org/10.3389/fnins.2014.00150

Hoff, K. A., & Bashir, M. (2015). Trust in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Joshi, K. D., & Nalwade, P. S. (2013). Modified k-means for better initial cluster centres. *International Journal of Computer Science and Mobile Computing, 2(7),* 219-223.

Kang, H. S., Lee, J. Y., Choi, S., Kim, H., Park, J. H., Son, J. Y., Kim, B. H., & Noh, S. do. (2016). Smart manufacturing: Past research, present findings, and future directions. *International Journal of Precision Engineering and Manufacturing-Green Technology*, *3*(1), 111–128. https://doi.org/10.1007/s40684-016-0015-5

Kozlowski, S. W. (2015). Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review, 5(4), 270-299.* https://doi.org/10.1177/2041386614533586

Kyriazos, T. (2018) Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology*, **9**, 2207-2230. doi: 10.4236/psych.2018.98126.

Lazzara, E. H., Benishek, L. E., Patzer, B., Gregory, M. E., Hughes, A. M., Heyne, K., ... & Schulman, C. (2015). Utilizing telemedicine in the trauma intensive care unit: does it impact teamwork?. *Telemedicine and e-Health, 21(8)*, 670-676.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust And Distrust: New Relationships and Realities. *Academy of Management Review*, *23*(3), 438–458. https://doi.org/10.5465/amr.1998.926620

Lowry, P. B., Schuetzler, R. M., Giboney, J. S., & Gregory, T. A. (2015). Is Trust Always Better than Distrust? The Potential Value of Distrust in Newer Virtual Teams Engaged in Short-Term Decision-Making. *Group Decision and Negotiation*, *24*(4), 723–752. https://doi.org/10.1007/s10726-014-9410-x

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A Temporally Based Framework and Taxonomy of Team Processes. *Academy of Management Review*, *26*(3). https://doi.org/10.5465/amr.2001.4845785

Mathieu, J. E., Gallagher, P. T., Domingo, M. A., & Klock, E. A. (2019). Embracing Complexity: Reviewing the Past Decade of Team Effectiveness Research. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*(1). https://doi.org/10.1146/annurev-orgpsych-012218-015106

Mathieu, J. E., Hollenbeck, J. R., van Knippenberg, D., & Ilgen, D. R. (2017). A century of work teams in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*(3), 452–467. https://doi.org/10.1037/apl0000128

Mathieu, J. E., & Rapp, T. L. (2009). Laying the foundation for successful team performance trajectories: The roles of team charters and performance strategies. *Journal of Applied Psychology*, *94*(1). https://doi.org/10.1037/a0013257

McNeese, N. J., Demir, M., Chiou, E. K., & Cooke, N. J. (2021). Trust and Team Performance in Human–Autonomy Teaming. *International Journal of Electronic Commerce*, *25*(1), 51–72. https://doi.org/10.1080/10864415.2021.1846854

Morrissette, A. M., & Kisamore, J. L. (2020). Trust and performance in business teams: a meta-analysis. *Team Performance Management: An International Journal*, *26*(5/6), 287–300. https://doi.org/10.1108/TPM-02-2020-0012

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, *115*(2). https://doi.org/10.1037/0033-2909.115.2.210

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*(2), 223–239. https://doi.org/10.1006/ijhc.1995.1042

National Academies of Sciences, E. and M. (2021). *Human-AI Teaming: State of the Art and Research Needs*. National Academies Press. https://doi.org/10.17226/26355

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2020). Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872082096086. https://doi.org/10.1177/0018720820960865

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 101–108. https://doi.org/10.1109/HRI.2016.7451740

Rupp, N. G. (2004). The Attributes of a Costly Recall: Evidence from the Automotive Industry. *Review of Industrial Organization*, *25*(1), 21–44. https://doi.org/10.1023/B:REIO.0000040514.22968.e1

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Ablex Publishing.

Feitosa, J., Grossman, R., Kramer, W. S., & Salas, E. (2020). Measuring team trust: A critical and meta-analytical review. *Journal of Organizational Behavior, 41*(5), 479-501.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Siebert-Evenstone, A., Michaelis, J. E., Shaffer, D. W., & Mutlu, B. (2021). *Safety First: Developing a Model of Expertise in Collaborative Robotics* (pp. 304–318). https://doi.org/10.1007/978-3-030-67788-6_21

Stajkovic, A. D., Lee, D., & Nyberg, A. J. (2009). Collective efficacy, group potency, and group performance: Meta-analyses of their relationships, and test of a mediation model. *Journal of Applied Psychology*, *94*(3). https://doi.org/10.1037/a0015659

Thayer, A. (2015). Trust trajectories as a function of violation type and repair efforts. *Ph.D. Thesis,* University of Central Florida Orlando, FL, USA,

Tag, B., van Berkel, N., Verma, S., Zhao, B. Z. H., Berkovsky, S., Kaafar, D., ... & Ohrimenko, O. (2023). DDoD: Dual Denial of Decision Attacks on Human-AI Teams. *IEEE Pervasive Computing, 22(1),* 77-84. doi: 10.1109/MPRV.2022.3218773.

Walliser, J. C., Mead, P. R., & Shaw, T. H. (2017). The Perception of Teamwork with an Autonomous Agent Enhances Affect and Performance Outcomes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 231–235. https://doi.org/10.1177/1541931213601541

Welfare, K. S., Hallowell, M. R., Shah, J. A., & Riek, L. D. (2019). Consider the Human Work Experience When Integrating Robotics in the Workplace. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 75–84. https://doi.org/10.1109/HRI.2019.8673139

Wildman, J. L., Lazzara, E. H., Shuffler, M. L., Fiore, S. M., Burke, C. S., Salas, E., & Garven, S. (2009). The development and validation of interpersonal trust and distrust scales. *Unpublished Manuscript, University of Central Florida, Orlando, FL*.

Wynne, K. T., & Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, *19*(3), 353–374. https://doi.org/10.1080/1463922X.2016.1260181

APPENDICES

Appendix A

Tables

Table 1

Descriptive Characteristics of Study Sample

	%	*N*
Gender		
Male	52.56%	41
Female	46.15%	36
Non-binary	1.28%	1
Race		
White/Caucasian	44.87%	35
Asian	35.90%	28
Black/African	16.67%	13
Native American	2.56%	2
Prefer not to disclose	3.84%	3
Asian	35.90%	28
Completed Education		
High school	17.95%	14
First year of college	19.23%	15
Second year of college	8.97%	7
Third year of college	5.13%	4
Fourth year of college	2.56%	2
Bachelor's degree	23.08%	18
Some graduate degree	6.41%	5
Master's degree	16.67%	13

Table 2.

Descriptive Statistics and Correlations across Variables.

	M	*SD*	1	2	3
1. Team Trust	5.54	.39	-		
2. Team Distrust	1.75	.45	-.520	-	
3. Build Time (minutes)	26.49	4.58	.185	.090	-
4. Accuracy (%)	68.79	24.12	-.312	.193	.145

Note: Vigilance was excluded due to binary classification of data.

Table 3.

Descriptive Statistics across Conditions.

	High Reliability					Moderate Reliability					Low Reliability				
	N	*M*	*SD*	Min	Max	*N*	*M*	*SD*	Min	Max	*N*	*M*	*SD*	Min	Max
Team Trust	13	5.76	.25	5.16	6.00	13	5.45	.40	4.69	5.97	13	5.42	.45	4.56	6.00
Team Distrust	13	1.74	.58	1.13	2.75	13	1.76	.36	1.13	2.38	13	1.76	.42	1.00	2.50
Build Time (minutes)	13	27.39	4.63	19.00	34.00	13	25.08	4.35	19.00	33.00	13	27.00	4.76	20.00	37.00
Accuracy (%)	13	53.73	21.38	25.00	100.00	13	67.42	22.02	47.06	100.00	13	85.22	19.08	50.00	100.00
Vigilance[1]	2(11)	--	--	--	--	6(7)	--	--	--	--	11(13)	--	--	--	--

[1]Vigilance is scored as a pass/pass/failure, represented here as number of teams who passed followed by number of teams who failed in parentheses: # Pass (# Fail).

Table 4.

Four Trust Profiles Identified via K-Means

	Mean Trust	Mean Distrust
Profile 1	5.87	1.23
Profile 2	5.67	1.98
Profile 3	5.13	1.80
Profile 4	4.98	2.49

Table 5.

Number of Teams Checking Kit Specifications by Condition

	[1]Teams (*N*)	Kit 1	Kit 2	[2]Kit 3	Kit 4	[3]Kit 5	Kit 6	[4]Kit 7	Kit 8	Kit 9
High Reliability	9	9	8	9	9	8	7	4	4	4
Moderate Reliability	12	12	11	11	11	10	9	9	10	9
Low Reliability	13	13	13	12	10	9	9	9	7	6

[1]Reflects number of teams in which data were available.

[2]Error introduced (low reliability only)

[3]Error introduced (low and moderate reliability)

[4]Error introduced (all reliability conditions)

Appendix B

Figures

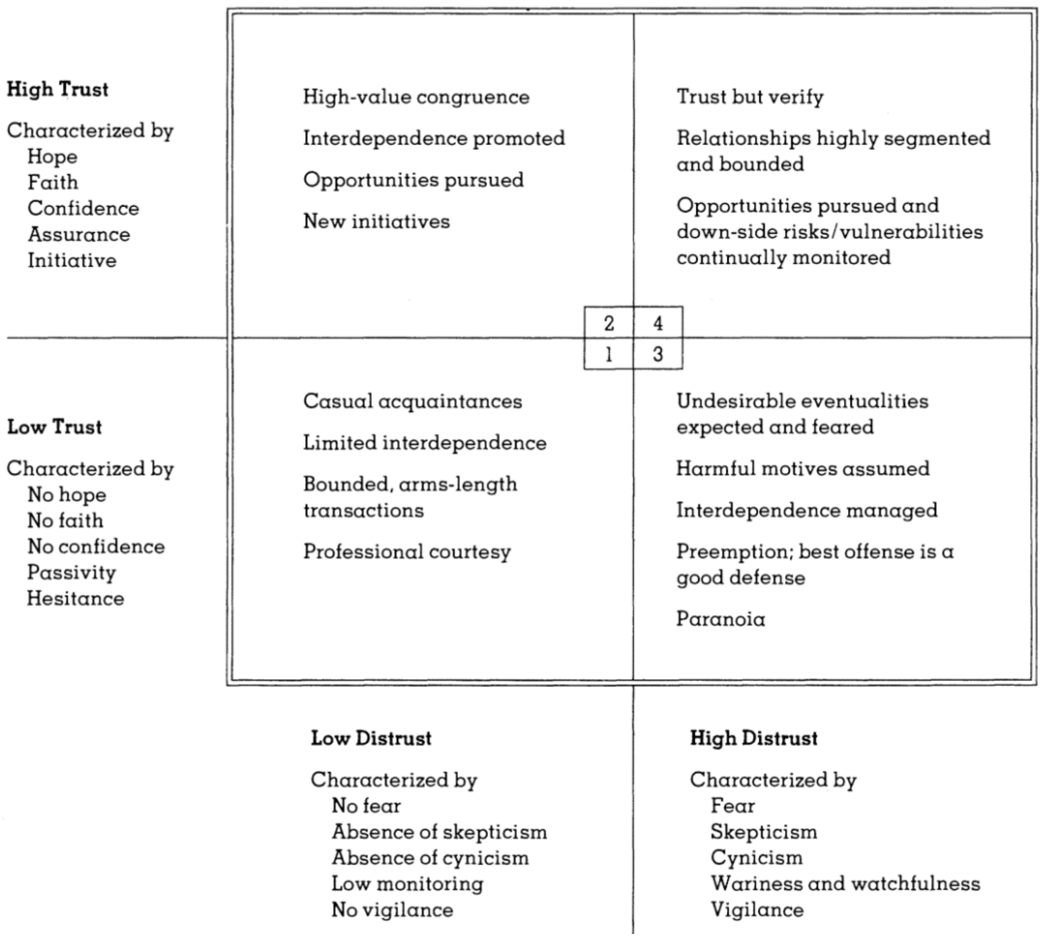

Figure 1: The Four Trust Profiles from Lewicki et al., 1998

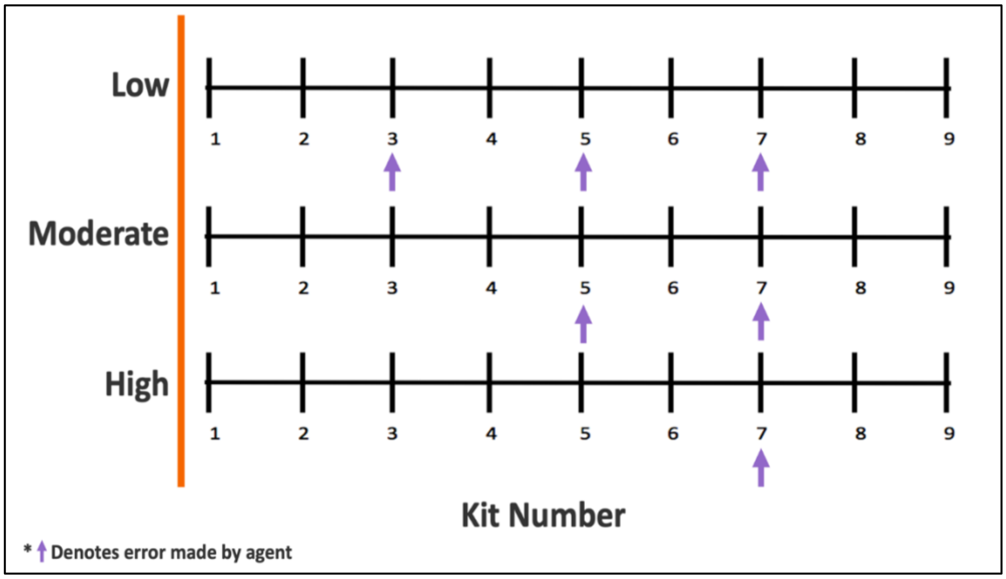

Figure 2: Depiction of error introduction across reliability manipulation

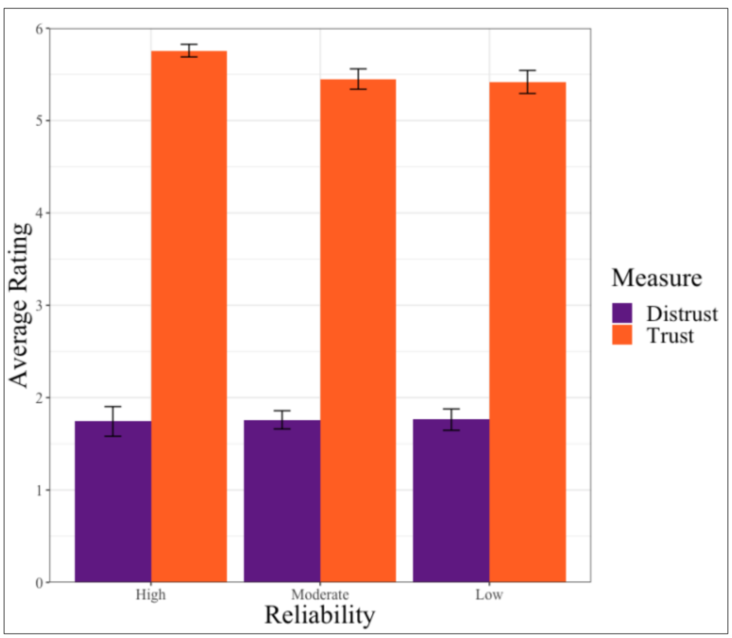

Figure 3: Average team trust and distrust across conditions

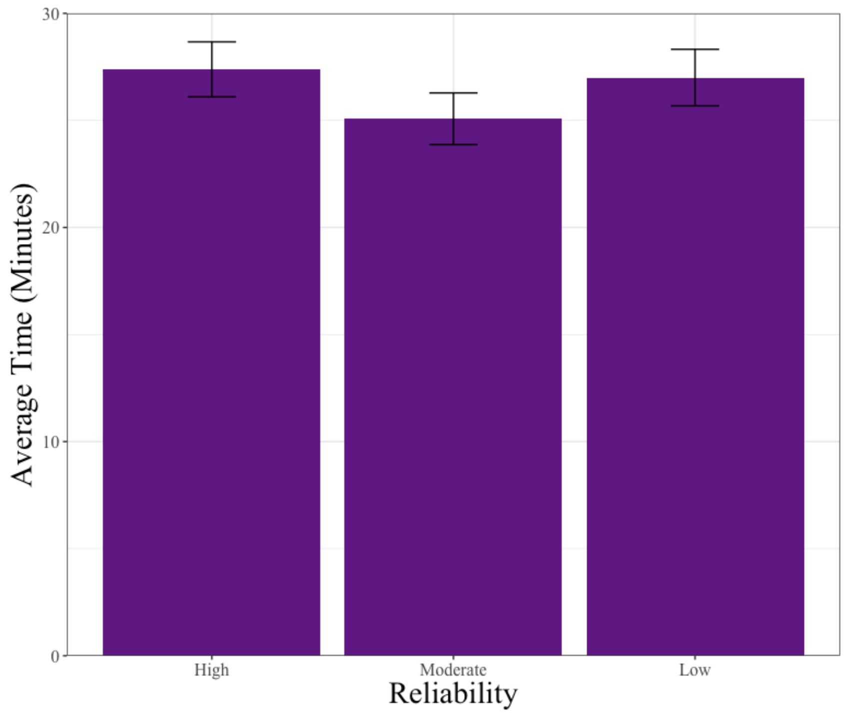

Figure 4: Average build time across conditions

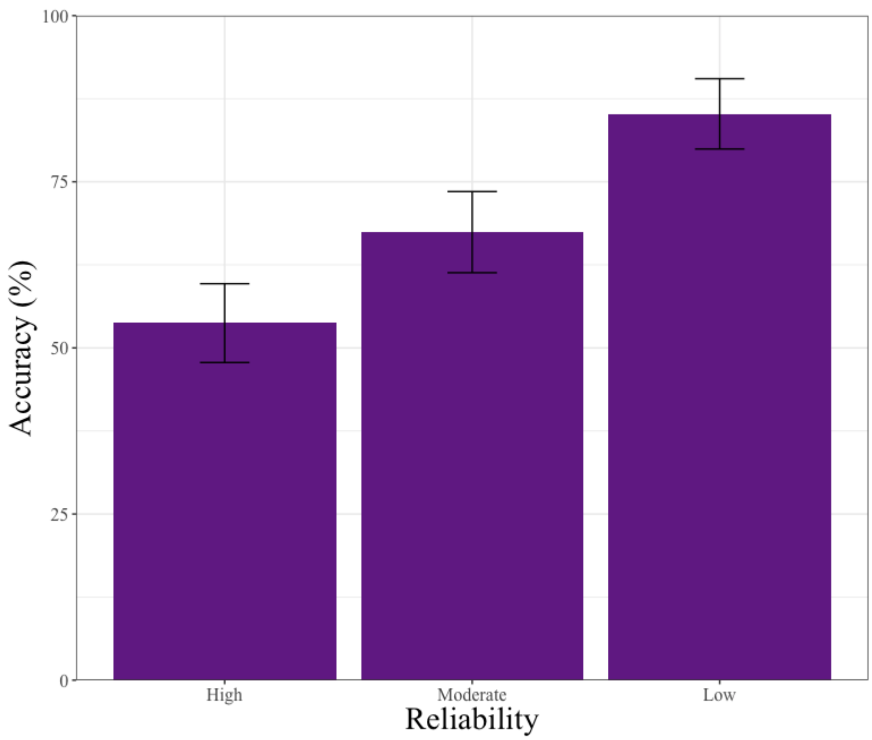

Figure 5: Average accuracy across conditions

APPENDIX C

Measure of Trust and Distrust

A trust and distrust scale developed by (Wildman et al., 2009) was used. Participants responded to the following items using a 6-point Likert scale (1 = not at all, 6 = very much so).

Thinking about your **human** teammate, to what extent did you feel:

1. Assured that this team member made intelligent decisions? *(Trust -Competence)*
2. Confident that this team member tried to do things that benefited the team? *(Trust – Intent)*
3. Afraid that this team member purposefully did something that wasn't helpful? *(Distrust – Intent)*
4. Faith that this team member could do the task at hand? *(Trust – Competence)*
5. Suspicious about this team member's reasons behind certain decisions? *(Distrust – Intent)*
6. Convinced that you could rely on this team member to try their hardest? *(Trust – Intent)*
7. Confident in this team member's ability to complete a task? *(Trust – Competence)*
8. Nervous that this team member would betray you? *(Distrust – Intent)*
9. Afraid that this team member would make a mistake? *(Distrust – Competence)*
10. Confident that this team member would do as they said? *(Trust – Intent)*
11. Positive that this team member would try and do what is best for the team? *(Trust – Intent)*

12. Compelled to keep tabs on this team member to be sure things got done? *(Distrust – Competence)*
13. Certain that this team member would perform well? *(Trust – Competence)*
14. Cautious about this team member's intentions for the team? *(Distrust – Intent)*
15. Paranoid that this team member would fail? *(Distrust – Competence)*
16. Worried that this team member would do something wrong? *(Distrust – Competence)*

Thinking about your **agent** teammate, to what extent did you feel:

1. Assured that this team member made intelligent decisions? *(Trust -Competence)*
2. Confident that this team member tried to do things that benefited the team? *(Trust – Intent)*
3. Afraid that this team member purposefully did something that wasn't helpful? *(Distrust – Intent)*
4. Faith that this team member could do the task at hand? *(Trust – Competence)*
5. Suspicious about this team member's reasons behind certain decisions? *(Distrust – Intent)*
6. Convinced that you could rely on this team member to try their hardest? *(Trust – Intent)*
7. Confident in this team member's ability to complete a task? *(Trust – Competence)*
8. Nervous that this team member would betray you? *(Distrust – Intent)*
9. Afraid that this team member would make a mistake? *(Distrust – Competence)*
10. Confident that this team member would do as they said? *(Trust – Intent)*

11. Positive that this team member would try and do what is best for the team? *(Trust – Intent)*
12. Compelled to keep tabs on this team member to be sure things got done? *(Distrust – Competence)*
13. Certain that this team member would perform well? *(Trust – Competence)*
14. Cautious about this team member's intentions for the team? *(Distrust – Intent)*
15. Paranoid that this team member would fail? *(Distrust – Competence)*
16. Worried that this team member would do something wrong? *(Distrust – Competence)*

APPENDIX D

Overview of Sub-Assembly Process

Participants were required to build a cart by constructing three sub-assembly kits and then compiling the sub-assembled kits into one cart. This was then repeated twice resulting in participants building three total carts. Kits contained parts for the upper frame, lower frame, and body. Samples of each are depicted below.

For the upper frame, participants received a kit with all necessary parts as well as a guidance sheet that listed all parts included and a reference photo (Figure 6).

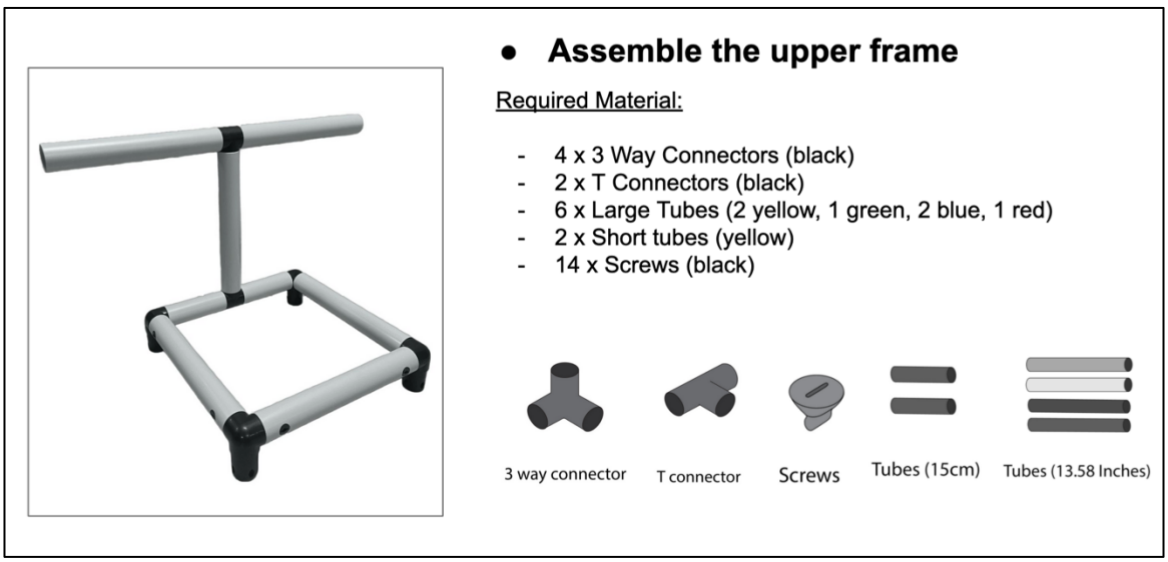

Figure 6: Sample upper frame guidance sheet given to participants.

For the lower frame, participants received a kit with all necessary parts as well as a guidance sheet that listed all parts included and a reference photo (Figure 7).

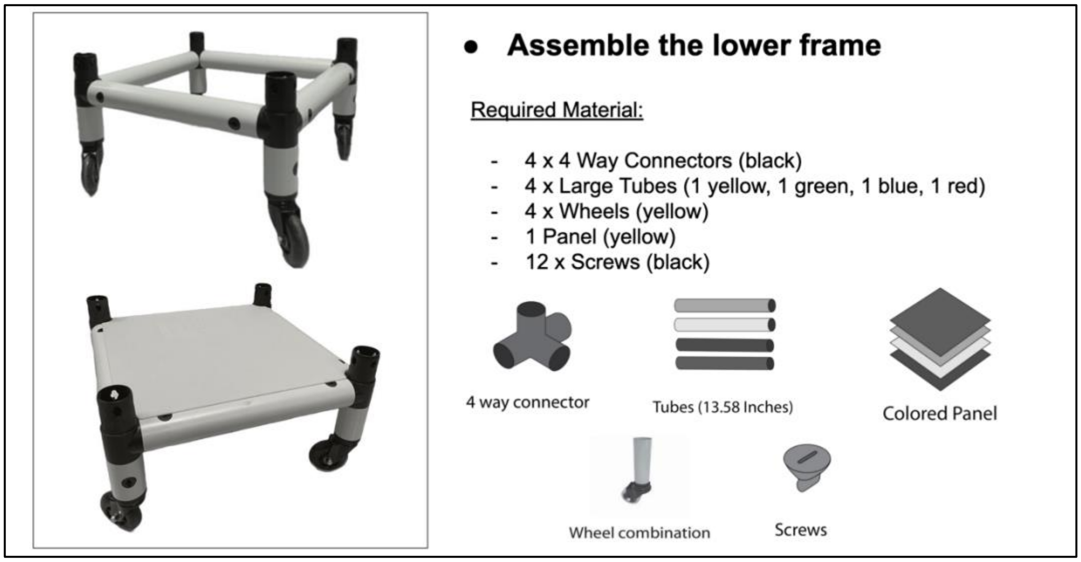

Figure 7: Sample lower frame guidance sheet given to participants.

For the body, participants received a kit with all necessary parts as well as a guidance sheet that listed all parts included and a reference photo (Figure 8).

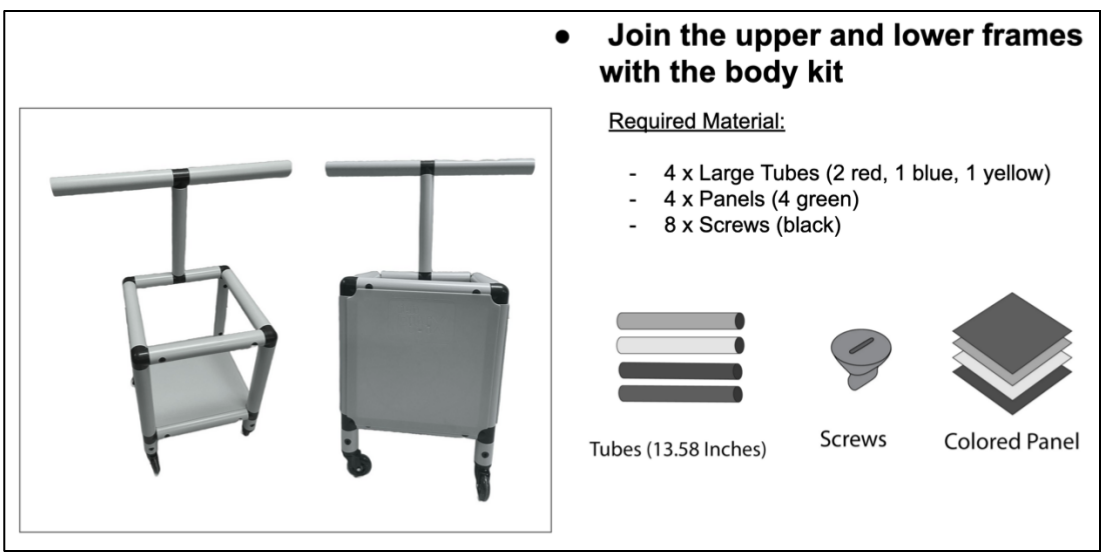

Figure 8: Sample body guidance sheet given to participants.