Florida Institute of Technology
Scholarship Repository @ Florida Tech

Theses and Dissertations

12-2020

# 1, 2, or 3 in a HAT? How a Human-Agent Team's Composition Affects Trust and Cooperation

Dan Manh Nguyen

Follow this and additional works at: https://repository.fit.edu/etd

Part of the Industrial and Organizational Psychology Commons

# 1, 2, or 3 in a HAT? How a Human-Agent Team's Composition Affects Trust and Cooperation

by

Dan Manh Nguyen

A thesis submitted to the College of Psychology and Liberal Arts of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Industrial/Organizational Psychology

Melbourne, Florida
December, 2020

We the undersigned committee hereby approve the attached thesis,
"1, 2, or 3 in a HAT? How a Human-Agent Team's Composition Affects Trust and Cooperation"
by
Dan Manh Nguyen

Jessica Wildman, Ph.D.
Associate Professor
Industrial/Organizational Psychology
Major Advisor

Meredith Carroll, Ph.D.
Associate Professor
Aviation Human Factors

Gary Burns, Ph.D.
Professor and Program Chair
Industrial/Organizational Psychology

Lisa Steelman, Ph.D.
Dean and Professor
College of Psychology and Liberal Arts

Abstract

Title: 1, 2, or 3 in a HAT? How a Human-Agent Team's Composition Affects Trust and Cooperation

Author: Dan Nguyen

Advisor: Jessica Wildman, Ph.D.

Modern advances in technology have enabled a collaborative relationship between man and machine. Many industries have adopted these human-agent teams, yet human perceptions about technology may prevent them from adopting a teammate mentality when interacting with agents. Although many studies have researched the issue, few have studied how the human to agent ratio within a team influences how the person intends to interact with their agent team members. Grounded in the theory of planned behavior (Azjen, 1985), this study elucidates how a team's composition affects the trust of human team members in human-agent teams and their subsequent intentions to work with their agent team members. Using a between-person experimental vignette methodology, 226 online participants were assigned to one of six vignette conditions in a survey which manipulates the composition of the hypothetical six-person team (agent majority, balanced, and human majority) and the role of the agent (leader or subordinate). Although few significant findings were produced, notable trends and study limitations are discussed to guide future research that examines the effect of team composition in human agent teams.

## Table of Contents

**List of Tables**

## List of Figures

## Dedication

I dedicate this thesis to my mother, who has provided unwavering support throughout my entire career. I also dedicate this thesis to my advisor, Jessica Wildman, for moving beyond an academic mentor to a personal inspiration who encouraged and supported me through the many setbacks posed by this year.

# Chapter 1: Introduction

One of humanity's trademark qualities is a capacity to develop tools in order to deal with the continuous challenges that present themselves in an evolving society. As humans continue to advance their understanding of the physical and social world in which they live, it has become increasingly clear how vast, deep, and ever-changing the mechanics of the everyday systems around them are. Like our predecessors before us, the same holds true in current society. Researchers are continuing to push and test theory in order to elevate our understanding of the interactions that occur around us, resulting in new developments in the many various fields of occupation as well as the emergence of new frontiers. However, unlike the challenges faced by our predecessors, the demands of the challenges in today's rapidly developing and complex society have begun to push us beyond our physical and mental abilities. As a result, humans have opted to reinforce themselves with technology (Orlikoswki, 2007). In doing so, humans have enabled themselves to tackle harder tasks and more tasks, as well as opened up new frontiers beyond previous human accessibility.

The need for technology itself is not novel. Throughout human history, we have always depended on our ability to understand our surroundings in order to fashion items that fulfill our needs. What has changed, however, is the nature of these needs. Due to changes caused by physical forces (e.g., weather and terrain shifts) and societal forces (e.g., collaboration and regulations), the nature of work today has become centered on progress (Volti, 2005). This shift from a survivalcentric society to a development-centric society has equivalently reflected a shift in technological use from sustenance to enhancement (Millar, Lockett, & Ladd, 2018). Indeed, the zeitgeist of the modern day is “smart” - people are using technology to make work easier, quicker, and plausible. Paralleling this change in utility, the mindset toward technology has gone from tool to teammate (Fiore & Wiltshire, 2016). While technology has historically been a means for problem-fixing, in recent times it has become a means for successful and efficient problem-solving. Injecting intelligence into technology has transcended its role from an object under our dominion to an entity with which we collaborate. Whereas a manufacturing company once employed machinery to simply accelerate production for its human workers, they have started providing robotic arms to work alongside humans during production (Cherubini et al., 2016). Where technology previously moved pieces from point A to point B, technology now moves pieces into positions for human workers to progress the build of a product (Michalos et al., 2010). As technology continues to “get smarter”, many industries beyond manufacturing continue to observe an illustrious partnership between man and machine such as the military (Jentsch, 2016), the medical field (Rastgarpour & Shanbehzadeh, 2011), emergency response/rescue units (Nourbakhsh et al., 2005), aviation (Kumar & Thakur, 2012), and sea/space exploration (Fong & Nourbakhsh, 2005).

The latter of these fields also demonstrates another important implication of evolving technologies: the emergence of new opportunities. Developments in aviation and sea technologies have enabled humans to explore new frontiers that were physically inaccessible, as well as opened up new opportunities in exploring these ventures such as improved reconnaissance and coordination (Olson et al, 2010; Ranganathan et al., 2010). More and more, people are equipping themselves with technological allies in order to access possibilities that were previously inconceivable. Space flight teams are often assisted by simulation and planning software that guide launches and path trajectories down to the minutiae that humans cannot (Marquez, Chang, Beard, Kim, & Karinksi, 2018). Military units survey new landscapes using aerial drones that convey perspectives and details humans cannot physically observe (Endsley, 2015).

It is clear then that technology inhabits an important part of our lives, especially in the current day and age in which there is more work to do that is harder in nature. Relationships with technology have gone from a toolbox utility to interdependent cooperation, leading to an increase in the use of human-agent teams (Shively et al., 2017). Yet in spite of their increasing popularity, their implementation has been far from flawless. One particular issue that these human-agent teams (HATs) commonly face is a lack of cooperation from human team members (Christoffersen & Wood, 2002; de Visser, Parasuraman, Freedy, Freedy, & Weltman, 2006; Leng, Li, & Jain, 2008; Steinfeld et al., 2006). The purpose of the current study is to examine how a tangible point of organizational intervention, the team's composition, affects a human team member's intentions to cooperate with their agent team member by elucidating the psychological process driving this relationship. Specifically, this study will use an experimental vignette methodology to manipulate the human-agent ratio in a team's composition and examine an individual human team member's subsequent trust and cooperative behavioral intention. By studying the varying trust that a human team member has toward their agent team members, and their team as a whole, between different team compositions, practitioners and scholars will be able to connect how staffing decisions impact a human team member's cooperation.

# Chapter 2: Literature Review

Like the actual application of HATs, research on HATs is spread out across multiple disciplines. This is due to the fact that various disciplines study the human component (e.g., psychology), the agent component (e.g., engineering paths and computer sciences), and the interaction between the two components (e.g., human factors). From the human component, human-agent research often draws from the traditional human teams literature for frameworks to examine the complex interactions between multiple individuals (e.g., IMOI models; Ilgen et al., 2005), as well as other lines of research on human affect (e.g., trust; Schaefer, Hill & Jentsch, 2018) and cognition (e.g., motivation, Jennings et al., 2014; decision-making, Parasuraman & Riley, 1997). From the agent component, human-agent research is informed by the fields of engineering and computer science to integrate the technological capabilities and design (e.g., agent architecture, Cayha & Giuliani, 2018; natural language processing, Runck, Manson, Shook, Gini, & Jordan 2019). As the field of human factors and ergonomics facilitates the marriage of these two areas of research arising from technological advances, research has shifted to studying the actual interaction of human and agent into a team context.

To understand the broader human perspective in a HAT, I first consult the vast literature on HATs and synthesize studies which investigates the interaction between both human and agent team members. Studying human-agent teams may take many forms as it primarily studies either or both the human and agent components of the team, but the interest of this study and review lies in the complex, dynamic interactions that unfold as a result of human psychology. Focusing on this human-agent interaction narrows the research of interest to research which connects to a human attribute and thus, purely agent-focused research (e.g., their technical design and development) are unincorporated into this review in favor of understanding the teamwork mechanics between human-agent team members. Both empirical and theoretical studies were included to provide larger insight into both what has been tested and what has been proposed.

*2. 1 Literature Search*

To capture the relevant studies from the various disciplines, the search strategy was broken up into two phases: a broad search and a narrow, discipline-targeted search. The initial broad search was intended to cast a larger net on potentially relevant research at large before identifying particular fields which often house research on human-agent interaction. To begin the broad search, comprehensive search engines (e.g., *Google Scholar* and the Florida Institute of Technology library's integrated database platform) were searched using the entry "*human agent teaming review".* From this, seven reviews were identified with four of these reviews broadly summarizing research on HATs (Chakraborti, Kambhampati, Schetuz, & Zhang, 2017; Chen & Barnes, 2014; Gao, 2013; Jennings et al., 2014) and three of these reviews integrating studies on specific topics within HAT research (Anjomshoae, Najjar., Calvaresi, & Främling, 2019; Schaefer, Hill, & Jentsch, 2018; Wright, Quinn, Chen, & Barnes, 2014). From these, an initial list of studies was created by extracting the citations from these reviews and removing duplicates. The titles and abstracts of these studies were then skimmed to manually identify if they fell within the scope of this review. After removing articles which were beyond the focus of this review, the remaining articles were read and labeled with tags indicating the topics they examined and the journal they were published in. A targeted search was then conducted by identifying the recurring disciplines from the journals the studies were published in and subsequently searching the major databases for those disciplines. To encompass the fields of psychology, computer science, engineering, and human factors, the databases *PsycInfo, ACM Digital Library, IEEE Xplore Digital Library,* and *Advanced Technologies & Aerospace* were searched using the keyword *human agent team**. New articles produced from the targeted search produced were additionally read and coded.

**Supplementary Search.** After reading the identified reviews and articles above, multiple new terms that were similar to human-agent teams arose which warranted a supplementary search. The keywords *human agent*, human autonomy*, human automation*, human machine*, and human robot** were each entered alone into the same comprehensive search engines and discipline specific databases from the literature search, and then again with every combination of keywords *team*, interaction, collective, collaboration, and integration*. The thesaurus function of the databases was also used to find other synonymous keywords that may have been missed, however no new terms were identified. After incorporating the results of the supplementary search, the final review consisted of 82 articles.

*2.2 Human-Agent Teaming*

Human-agent teaming focuses on groups in which humans and intelligent technologies work interdependently on tasks to achieve an objective (Chen & Barnes, 2014; Russell & Norvig, 2016). Human-agent teams (HATs) are identified using similar criteria to defining a team (i.e., there are two or more members whose tasks are interdependent and work towards shared goals; Salas, Rosen, Burke, & Goodwin, 2009) with the added distinction of having an autonomous machine (i.e., an agent) as one or more of its constituent members. These criteria are relatively straightforward, however as seen from the literature search process, there is abundant terminology for referring to some form of human-technology interaction. Although nuances exist between these many terms (e.g., automation, machine, and agent), a commonly accepted definition from Russell and Norvig (2009) states that agents must be autonomous, observe their environment, and act upon the environment.

Although relatively nascent in use compared to traditional human only teams, several industries have begun implementing HATs to increase efficiency and access new opportunities. In the military, combat units have been supplemented with artificial intelligence to guide attack drones and augment human precision and targeting (Endsley, 2015). In emergency response teams, search & rescue robots have been deployed to access hazardous environments unsuitable to humans in order to save lives (Nourbakhsh et al., 2005). Specialty manufacturers (e.g., aerial, naval, and other large custom machinery producers) have begun implementing intelligent robots in conjunction with technicians to craft complex mechanical components that require simultaneous actions to consolidate multiple steps (e.g., a robotic arm fixes the next piece in the process in place for a mechanic to weld; Valente, 2016). Additionally, health-related treatment and prevention techniques have seen improved healthcare efficiency from the use of machine-learning algorithms used in tandem with medical practitioners (Rastgarpour & Shanbehzadeh, 2011). Clearly, the partnership between humans and agents demonstrates impressive potential for effective performance. What is less clear is understanding why and when this potential is fulfilled, compared to instances when human-agent teamwork breaks down. To address this, the following sections of the review are organized using an input-mediator-output framework similar to the IMOI model (Ilgen et al., 2005) to summarize relevant factors as they correspond to these phases of a team's existence (see Figure 1 for a meta-model of the literature).

### *2.3 Inputs*

Before the members of a group even interact, multiple existing factors within the team will influence its future. These factors that precede the interactions within a group are referred to as the inputs to a team (McGrath, 1984). These inputs are often properties of both the individuals within a team and the team itself at large. Team process models thus often distinguish individual level inputs from team level inputs and study them differently in line with multilevel theories

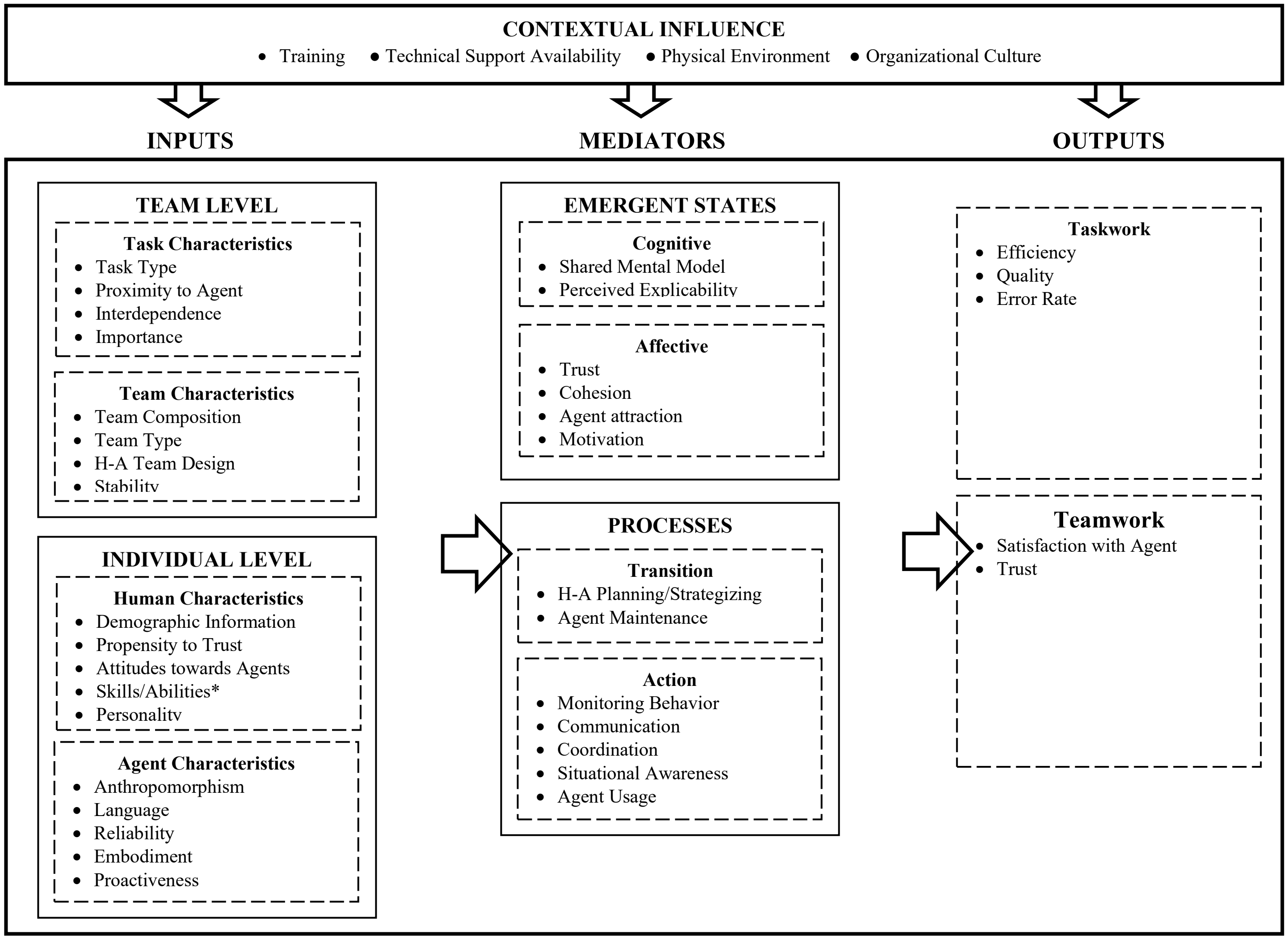


Figure 1. *IMOI-Based Framework of Human-Agent Research*

(Kozlowski & Klein, 2000). The addition of an agent in HATs changes both team level properties and the individual level properties when compared to traditional human teams and thus shifts the focus of interests when it comes to pinpointing the inputs with the strongest implications for the team's future. Concerning the team at large, research on inputs to HATs have identified a variety of characteristics that impact subsequent mediators and outputs. As they relate to the human interaction, these team level characteristics can be classified into two categories: task characteristics and team characteristics.

As the purpose of a team is to work together towards a common goal (Salas, Rosen, Burke, & Goodwin, 2009), its members' tasks are housed within a larger team objective. The tasks that are assigned within a HAT must consider the additional implications of how human team members respond to certain tasks being assigned to agent team members. Research on agent task characteristics have shown that agents are thought of as less capable (Gombolay, Huang, & Shah, 2015) and overridden more (Parasuraman & Riley, 1997) when they are responsible for tasks that are perceived to be critical to shaping a team's outcome. Similarly, research on complacency has found that people have a tendency to trust agent team members with objective and technical tasks (e.g., computation, processing information), but trust them less with subjective tasks (e.g., decision making; Gombolay et al., 2015). Yet, this tendency to undervalue agent team members concerning higher stake tasks is not always observed when a human team member's task is linked to an agent's task. When tasks are highly interdependent, human team members are more likely to be cooperative and perceive the computer to be similar to themselves (Nass, Fogg, & Moon, 1996). This phenomenon of increased performance when bringing human and agent team members closer is paralleled in research on task proximity (Gabler, Stahl, Huber, Oguz, & Wolherr, 2017). Gabler and colleagues (2017) found that an agent's decision-making ability was respected and acknowledged in HATs where human team members worked on an assembly task before handing it over to an agent team member for placement in a warehouse.

Inputs referring to team characteristics may be thought of as the available resources that make up a team (Kozlowski & Bell, 2003) and pinpoint several variables unique to HATs which influence the team's interactions and outcomes. Like their organizational human counterparts, a HAT may be made up of a variety of members who possess diverse characteristics. However, by virtue of its definition, a HAT's composition has the added criterion of containing an agent team member. Although this is an obvious statement, what is less apparent is the implications it has for the team's dynamics. Research on the different components of a HAT's composition has empirically studied human-to-agent ratios (Burke & Murphy, 2004; Murphy, Burke, Barnes, & Jentsch, 2010), team size (Mendonça, Brooks, & Grabowski, 2014), subgroup differences (Robert & You, 2015), and cognitive diversity (Sauer, Felsing, Ranke, & Rüttinger, 2006) , as well as additional theoretical consideration (You & Roberts, 2018). Research on the human-agent ratio of a team has shown that teams composed of two human team members and one agent team member yield the best results (Burke & Murphy, 2004; Murphy, Burke, Barnes, & Jentsch, 2010), however these findings are derived from analyses of archival field data and limit the causal inferences that can be made about the effectiveness of this ratio, especially as it relates to predicting any other team processes and emergent states. Although these studies offer important evidence into understanding the effect of human-agent composition, HAT researchers have called for more theoretically driven empirical studies (Teo, Wohleber, Lin, & Reinerman-Jones, 2017).

The nature of the team, such as its context and operating environment, have been found to necessitate certain collaboration and coordination protocols (Neef, 2006). For example, teams in high octane, adaptive environments such as surgical teams and rescue teams may lean towards teamwork oriented collaboration with a human in command to effectively respond to rapid changes that may occur (Nourbakhsh et al., 2005), while procedural teams such as manufacturing teams may lean towards mixed initiative teams where both the human and agent team members share control of the team's tasks (Owan, Garbini, & Devasia, 2017). In a similar vein, the design of the team (i.e., human-agent roles, communication channels) will also dictate its operations (Chen & Barnes, 2014). Research into human-agent roles have found that teams which appropriately allocate the strengths of its human and agent team members will perform better in dynamic, constantly changing missions (Bradshaw et al., 2008; Goodrich & Schultz, 2008). It is preferable for teams in these dynamic environments to remain stable in its operations, with the same team procedures and members in place for multiple missions, to fully leverage the resulting team dynamics (Demir, Cooke, & Amazeen, 2018).

The individual level inputs to a HAT may refer to either the characteristics of a human team member and the characteristics of an agent team member. Beginning with characteristics of human team members, research on traditional human teams has identified a large number of individual differences which influence a team's processes and outcomes and generally examine the surface- and deep-level traits that its members possess (Lyons & Guznov, 2019). The following research on human characteristics in a HAT do not make such a distinction and rather broadly identify key individual differences related to technology which influence a person's relationship with their agent team members, such as demographics (gender, age), predispositions, and self-efficacy. Demographic characteristics are a staple to studies in any social domain, however research on classroom technologies has specifically found that men (Dunne, 1998) and younger individuals (Czaja & Sharit, 1998) tend to be more comfortable with technology. The preconceived attitudes that one carries also influences how they will act towards the targets of these attitudes. The trust an individual in a HAT has in their teammates depends on their propensity to trust (Schaefer, Hill, & Jentsh, 2018; Singh, Molloy, & Parasuraman, 1993), and propensity to trust automation, which has been argued to be different from propensity to trust humans (Nickerson & Reilly, 2004). More specifically, a person also holds attitudes towards technology which may increase or decrease the initial trust they have in their agent team members (Backonja et al., 2018; Merritt & Ilgen, 2005). A person's previous experiences with technology often shapes this, as well as their confidence to successfully operate and interact with technology (de Vries, Midden, & Bouwhuis, 2003). Although many other individual differences studied from traditional human teams have been empirically examined within HATs as well (enough to deserve its own review), the three characteristics studied above have shown to be important factors which relate a person's individual differences to the workings of a HAT.

As agents are the defining piece which distinguishes a HAT from traditional human teams, research on the design and characteristics of an agent are abundant and stem from many fields. Of these many design considerations, several are more prominent when trying to understand the human perspective. First, because the agent is a team member, people pull from their interactions with the agent to mold their thoughts and attitudes similar to the beginning of any relationship. Research has shown that a person is more likely to trust an agent when it expresses more human-like qualities, and that negative behaviors resulting from low trust such as misuse are reduced when the agent possessed more anthropomorphic traits (de Visser et al., 2016; Parasuraman & Riley, 1997). This line of research has drawn from research on human-animal teams to identify the psychological components of perceiving human characteristics, such as playful behavior and identifiable emotions from facial expressions (Billings et al., 2012; Philips et al., 2016). Interactions with agents are also influenced by the agent's appearance beyond invoking varying senses of familiarity or discomfort. The embodiment of the agent acts as a direct medium through which human senses interact with the agent, consequently affecting how people interpret intentions and desires from an agent's actions (Stowers et al., 2016). However, research is not consistent in regards to the importance of tangibility, as agents manifest in studies in varying forms with varying results (e.g., physical robots and software rooted within a larger system; Parasuraman & Miller, 2004). It is also interesting to note that the appearance of physical robots triggers gender role stereotypes. Eyssel & Hegel (2012) found that when a robot appeared masculine, people were more likely to assign them to stereotypically male roles such as maintenance and repair work, whereas robots which appeared feminine were commonly given female tasks such as caregiving and service work.

Research from technical fields (e.g., computer science) has also examined the importance of more agent-centric characteristics such as reliability. An agent's reliability is often studied as it predicts the trust that a human team member will have towards the agent. Numerous studies have examined this, and found positive correlations between reliability and a human team member's initial trust (Fan et al., 2008; Hancock et al., 2011, Chiou & Lee, 2016). However, when an agent performs at a lower level than expected from its level of reliability, a human team member's trust in the agent will decrease more than it would towards another human team member who dropped in performance (de Visser, Pak, & Shaw, 2018).

*2.4 Mediators*

Mediators are the means by which a team turns its inputs into outcomes and refers to any factor that drives this conversion process (Ilgen et al., 2005). Whereas the IPO model of team systems labeled these factors as processes (Hackman & Morris, 1975), the IMOI framework extends this term to accompany emergent states as well to capture important, non-behavioral mechanisms which also push a team towards its output. Compared to traditional human teams, research on HATs has focused on how particular emergent states and processes occur differently than they do in human teams to underscore the importance of different mediators. Although HATs add two new interaction relationships (human-agent and agent-agent) to the human-human interaction studied in organizational team research, this review only focuses on research on human-agent interaction to shed light on the human perspective.

**Emergent States.** Emergent states that arise from human-agent interactions provide insight into understanding when and how human team members act differently towards their agent team members because they reveal certain psychological mindsets that people develop as they interact with agents. Research on cognitive emergent states often studies how human and agent team members share understanding and perceive the task environment around them (Goodrich & Yi, 2013; Nikolaidis & Shah, 2012; Scheutz, DeLoach, & Adams, 2017). Research on team cognition within HATs is a topic of popular interest, with articles in the human-agent interaction stream focusing on how shared mental models are developed across human team members and agent team members (Fan & Yen, 2010; Fan et al., 2017; Goodrich & Yi, 2013; Nikolaidis & Shah, 2012; Perelman, Evans III, & Schaefer, 2017; Scheutz, DeLoach, & Adams, 2017; Talamadupula, Briggs, Chakraborti, Scheutz, & Kambhampati, 2014; Yen et al., 2006). Unsurprisingly, a HAT's shared mental model is more similar when its agent's technology enables them to predict and articulate the needs of their human team members (Fan et al., 2017).

What is less obvious, however, is how the human pieces to this equation contribute to this increased convergence in the team's shared mental model. Fan and Yen (2010) found that human team members were able to invest more of their cognitive energy into their task instead when agents had these improved planning and communication abilities. In a spatial navigation experiment, Perelman, Evans III, and Schaefer (2017) further found that human team members were more likely to adapt their mental model to match an agent's input if the agent showed similar mental models of the environment and task by suggesting routes similar to the human team member. Research on the topic of explicability intersects with this phenomenon to further explain how an agent's suggestion influences a person's willingness to change their own mental model. Specifically, Meszaros, Le Vie, & Allen (2018) found that when their agent generated a plan similar to a human team member's plan, that person will be more likely to understand or explain an agent's decision that differs from their own within a moderate margin (roughly 14% difference). Although these studies on shared mental models in HATs often found increased team performance as well, research at large has not examined the accuracy of these converging mental models.

Affective emergent states play a vital role in understanding why HATs run into the various collaboration issues between its human team members and agent team members. Several affective states have been studied including team cohesion (Zieba, Polet, Vanderhaegen, & Debernard, 2010), agent attraction (Prada & Paiva, 2009), and motivation (Krippendorff, 2004), but none more than trust (Schaefer, Hill, & Jentsch, 2018). A human team member's trust in their agent team members (popularly abbreviated as H-A trust; Chen & Barnes, 2014) has received special attention due to its proximal influence on many team processes such as communication (Demir, McNeese, & Cooke, 2016), coordination (Demir et al., 2018; Gombolay, Huang, & Shah, 2015; Talamadupula, Kambhampati, Schermerhorn, Benton, & Scheutz, 2011), and proper usage (Parasuraman & Riley, 1997). H-A trust has been defined by Lee and See (2004) as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (p. 54). The research on human-agent trust is vast and has multiple stand-alone reviews (Chen & Barnes, 2014; Lee & See, 2004; Madhavan & Wiegmann, 2007; Schaefer, Hill & Jentsch, 2018) and a meta-analysis (Hancock, et al., 2011) which explore the many relationships between H-A trust and its antecedents and consequences.

From these reviews, several prominent findings emerge which highlight important issues in human-agent teaming from the human perspectives. First, human team members may initially trust the ability of an agent more than another human team member, but they are more sensitive to any actions which violate this trust (deVisser, Pak, & Shaw, 2018; Jian, Bisantz, & Drury, 2000). Further research on H-A trust violation elaborates on the slope of the subsequent drop in trust (deVisser, Pak, & Shaw, 2018), but the sensitive nature of human trust towards agents has caught the focus of researchers. A meta-analysis by Hancock and colleagues (2011) found that factors related to the agent's performance (e.g., reliability, failure rates) were found to be better predictors of trust than factors related to the agent's traits (e.g., anthropomorphism, personality), which may suggest that a human's trust towards an agent revolves around tasks. Dzindolet and colleagues (2001) also found that environmental factors related to the tasks connecting the human and agent played an important role in understanding when a person's perceived reliability of an agent differs from their perceived reliability of a human, and specifically found that H-A trust was lower when the task was perceived to be riskier or more complex.

Beyond these findings concerning human perception in H-A trust, it is worth mentioning that research on H-A trust has also examined other parts of the trust process such as trust repair both empirically (de Visser, Pak, & Shaw, 2018) and theoretically (Marinaccio, Kohn, Parasuraman, & de Visser, 2015). In sum, the overall guiding takeaway for this study from the research on H-A trust is that trust is a key affective state which is not only related to a host of outcomes, but other team processes as well.

**Processes.** Like all teams, HATs transform their inputs into outputs through the activities of its team members (McGrath, 1984; You & Roberts, 2018). Team processes are thought to be temporally based as a team's actions sequentially unfold over time (Marks, Mathieu, & Zaccaro, 2001). As its members engage in activities to both prepare for a task and execute the task, any given team process occurs in either a transition phase or an action phase (Marks, Mathieu, & Zaccaro, 2001). Transition behaviors in HATs that have been studied include a variety of agent-related preparations such as human-agent planning and strategizing (Van Diggelen, Neerincx, Peeters, & Schraagen, 2018) and agent maintenance (Hobbs, 2008). Action behaviors studied in HATs are numerous as well, and include topics such as monitoring behaviors (Kaminka, Pynadath, & Tambe, 2002), communication (Demir, McNeese, & Cooke, 2017; Tweedale et al., 2008), coordination (Shah & Breazel, 2010), situational awareness (Chen et al., 2018), and agent usage (Parasuraman & Riley, 1997). While each of these action behaviors have been identified as relevant to the operation of a HAT, research on the human usage of agents carries significant weight in the discussion of human-agent interaction, as the inappropriate use of an agent is commonly cited as the reason for collaborative breakdowns in HATs (Christoffersen & Wood, 2002; de Visser, Parasuraman, Freedy, Freedy, & Weltman, 2006; Leng, Li, & Jain, 2008; Steinfeld et al., 2006). Parasuraman and Riley (1997) identified 4 types of usage behaviors: use, misuse, disuse, and abuse. Under this categorization, use refers to behaviors where a human correctly employs a machine's assistance. Misuse refers to behaviors in which automation is allowed to act beyond its intended boundary, and can be thought of as an action that over-relies on automation. Meyer and Lee (2013) further explained that misuse can be distinguished into actions of compliance, in which a human simply accepts input from automation without considering its accuracy, and reliance, in which a human actively seeks out the automations assistance beyond its intended use. Further research from Wickens, Clegg, Vieane, & Sebok (2015) identified the sources of reliance (e.g., complacency) and compliance (e.g., automation bias). Disuse is seen as the opposite of misuse and refers to any ignoring behaviors which underutilize automation as intended. Lastly, abuse refers to human behaviors which bypass or defeat the purpose of the automation's implementation without considering the possible repercussions to the humans involved. Examples of abuse include intervening (e.g., overriding an automation to do its task for it; Leng, Li, & Jain, 2008) and disabling (e.g., shutting down an automation; Lee, 2006). It is worth noting that research on failures in human-agent collaboration often cites these particular examples of intervention and disabling as a recurring problem for HATs (Battiste et al., 2018; Beck, Dzindolet, & Pierce, 2007; Christoffersen & Wood, 2002).

### *2.5 Outputs*

The results of interest from a HAT are often the same taskwork and teamwork outcomes studied within traditional human teams (Hancock et al., 2011). However, several specific differences are worth highlighting. Regarding taskwork outcomes, objective metrics (e.g., efficiency, quality, error rate) for performance at the team level may include evaluations of human-agent interactions (Gouman et al., 2010; McNeese et al., 2018). Not many additional teamwork outcomes are novel to HAT research, however the measurement of these outcomes differ from traditional human teams due to translating scores between human and agent team members (e.g., mental models, Fan et al., 2017; information sharing; Demir et al., 2015). Additional research has also examined further attitudinal outcomes such as a human team member's satisfaction with their agent (Yan et al., 2013) and the acceptance of their agent (Demir, Cooke, & Amazeen, 2018).

*2.6 Contextual Influences*

Like any team, HATs are also embedded in a larger context which constantly influences its inputs, mediators, and outputs. These are especially important to the members of a HAT as the higher-level factors (e.g., organizational resources and protocol), play an important role in shaping the human-agent interaction within the team at any time. Research has identified that training (de Visser et al., 2006; Nikolaidis & Shah, 2013), technical support (You & Roberts, 2018), physical environmental factors (Hancock et al., 2011), and the organizational culture surrounding automation (Evers, Maldanado, Brodecki, & Hinds, 2008; Wang et al., 2010) may be particularly influential factors in a HAT that affect levels of a team's input, mediators, or outputs, and their relationships to one another.

# Chapter 3: Hypothesis Development

Although human-agent teams have been implemented across many industries, their use has not been without issue. Because of the long-standing status of machines as tools rather than teammates, as well as the agent-human dichotomy inherent in human-agent teams, humans do not tend to interact with machines in the same way they do with other humans (Bradshaw et al., 2008). Namely, people are less likely to cooperate with agents (Christoffersen & Wood, 2002; Gombolay et al., 2015). This breakdown in teamwork has been a persistent challenge in human-agent teams that has been documented in the literature for the past two decades (Battiste et al., 2018; Beck, Dzindolet, & Pierce, 2007; Christoffersen & Wood, 2002; Schaeffer, Hill, & Jentsch, 2018).

The purpose of the current study is to contribute to research on the teaming problem in human-agent teams by experimentally examining how one input to HATs, the team's ratio of humans to agents, influences a human's attitudes and behavioral intentions. Although some prior studies have begun studying HAT composition, and there may be skepticism about the usefulness of HAT composition research since the team's membership may be dictated by more pressing conditions (i.e., task requirements, available resources), this study has merits in addressing the teaming problem in HATs. Specifically, this study may provide novel insight into how team composition influences an individual's intent to act cooperatively through a theory-driven experimental design elucidating the thoughts and attitudes that human team members experience prior to making a decision. Whereas prior research primarily focuses on the attitudes of human team members towards their agent team members, this study additionally focuses on how team composition influences the attitudes of human team members towards the entire team.

Although a limited number of empirical studies on human-agent team composition exists, they are all data-driven approaches derived from observations of HATs operating in the same high-stake environment. For example, a field study using data from disaster response training by Burke and Murphy (2004) suggests that a 2:1 human-to-robot ratio yielded the best results, as teams with two human operators performed better than teams with one human operator due to improved situational awareness during the rescue task. A follow-up study by Murphy, Burke, Barnes, and Jentsch (2010) corroborated this 2 human to 1 robot ratio for being the ideal team composition, however this recommendation was specifically derived from and intended for high intensity environments such as search and rescue teams and military combat teams. This ratio was suggested based on the fact that introducing autonomous robots splits the attentional demands of human team members and poses increased safety risks that could jeopardize human lives. Although these studies offer valuable insight into developing research on HAT composition, neither of these two studies developed a-priori hypotheses to test and confirm the presence of a natural phenomenon. While these data-driven studies have an important place in making sense of natural events and highlighting important concerns, the current study will add to these field studies by testing theoretically driven hypotheses related to a HAT's composition in a controlled, experimental study.

In some cases, the composition of a human-agent team is determined by pre-existing factors such as the availability of resources (i.e., funding, number of available units) or the nature of the team's task (i.e., the number of spots, the nature/danger of certain actions). Indeed, an empirical study based on data from the 9/11 attack on the World Trade center by Casper & Murphy (2003) observed that different ratios exist for different purposes that correspond to the purpose of a team. A transportation ratio refers to the number of humans required to incorporate the agent (i.e., the number of people needed to enable the agent's device or literally lift the agent's device to the task location), while an operation ratio refers to the minimum number of humans and agents that are each needed to carry out the task. However, not all HATs are bound by a condition which restricts their composition (i.e., management/task coordination software; Keen, 1980). A HAT's composition may be restricted by resources, but that does not mean the resulting composition is ideal. Decision-makers with the flexibility to staff HATs would thus benefit from research exploring the ideal human-to-agent ratio in a team to consider whether or not it is worthwhile to invest more resources to achieve that ideal composition.

In sum, although this study acknowledges the research on HAT composition before it and acknowledges that there are situations in which the HAT's composition is restrained by more important circumstances, this study contributes the theory-driven research on HAT composition that is missing in the literature to provide suggestions for HATs which have the liberty to make staffing decisions. Using Azjen's theory of planned behavior (1985), this study will examine how a HAT's composition will influence an individual's intention to cooperatively behave with an agent as explained by their attitudes and cognition.

### *3.1 Theory of Planned Behavior*

To understand the human perspective of this issue, this study is framed using the theory of planned behavior (Azjen, 1985) which states that the subjective norms surrounding an action, an individual's perceived control over the matter, and an individual's attitudes towards an action will determine an individual's intent to carry out a certain action before actually engaging in the action. The theory of planned behavior thus emphasizes that a behavioral intention precludes an actual behavior itself, and reveals three mechanisms (attitudes, perceived control, and subjective norms) that predict this intention. Inputs which affect these mechanisms would thus influence a person's behavioral intention and ultimately, their behavior. This study focuses on the attitudinal component to leverage prior research highlighting the importance of trust, and posits that team composition, as conceptualized through majority/minority categories, may be one such targetable input for HATs that would affect the perceptions and intentions of a human team member.

### *3.2 Behavioral Intention*

Returning to the teaming problem in HATs, it has often been observed that human team members in a HAT will not cooperate with their agent team members

(Battiste et al., 2018; Beck, Dzindolet, & Pierce, 2007; Christoffersen & Wood, 2002; Schaeffer, Hill, & Jentsch, 2018). This decision to not work with their agent as a teammate can also vary in extremity and range from a lack of acknowledgement (i.e., ignoring an agent's input/action) to counterproductive behavior (i.e., acting contrary to the agents input/action, overriding the agent to redo its action). However, as research on agent use has shown, a human team member may also comply with their agent team member and accept its input (Parasuraman & Riley, 1997). Considering both ends of the spectrum, this range of behavior is representative of a larger continuum of cooperative behavior.

Complementary to the literature, which has established that human team members do engage in a range of cooperative behaviors, this study examines the intention to engage in these behaviors as opposed to the actual display of these behaviors themselves. While the behaviors that a human team member exhibits are often the outcome of interest, since behaviors are ultimately tied to consequences, understanding the behavioral intention that occurs prior to the behavioral can better elucidate the actual attitudes and thoughts that a person experiences leading up to their behaviors (Sheeran, 2002). Taken one step further, pinpointing factors that influence these attitudes and thoughts provides a tangible point of intervention to enact change. Although behaviors themselves are an important outcome of ultimate interest, the current COVID-19 pandemic has limited the ability to conduct the in-person laboratory research needed to observe actual behavior.

Within this study, cooperative behavioral intention refers to the degree that an individual human team member plans to productively act in accordance with an agent team member's action. A human team member exhibits higher cooperative behavioral intentions when they indicate that they plan to accept their agent team member's input and act accordingly. A human team member exhibiting low cooperative behavioral intentions would indicate that they plan to reject their agent team member's input and act counterproductively to their agent team member. A human team member exhibiting moderate levels of cooperative behavioral intention would indicate that they will ignore their agent team member.

### *3.3 Team Composition*

Team composition is a broad term which has traditionally referred to how attributes of a team are configured (Levine & Moreland, 1990). Research on team composition in human teams have primarily examined the capabilities of team members (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995) as well as the effect of surface-level and deep-level differences between team members (Bell et al., 2018). While the KSAO approach may not provide novel insight into the interaction between human and agent given the standardized quality of agents (i.e., agents are implemented when they are shown to be highly reliable and useful), research into the effect of surface level composition has revealed several findings that may be important for understanding human-agent team composition. For example, the perception of the surface-level characteristics in a team precedes its effects, meaning that people in a team vary in the extent that they are attuned to differences (Harrison, Price, Gavin, & Florey, 2002). Research has also shown that members of a team will have lower levels of satisfaction when they perceive a division based on some characteristic (Jehn & Bezrukova, 2010). This division has also been shown to be related to negative team behaviors such as less information sharing between the two groups in the team (Lau & Murnighan, 2005). Taken together, these findings suggest that human team members pay attention to differences in a team along a dividing characteristic that influences how they will interact with their team members. This directs focus within team composition to the human-agent division in HATs, and in this study specifically, the number of agent team members versus the number of human team members. While team composition can refer to the make-up of a team' in terms of any attribute, team composition in general is empirically understudied in HATs. Given these two considerations, and as a first step into the foray of HAT composition research, this study focuses on the most basic type of HAT composition: the ratio of humans to agents. That is, the ratio of agent team members to human team members is hypothesized to predict the attitudes and thoughts that human team members will undergo to direct their behavioral intention.

Although a continuous metric assessing the ratio of agent team members to human team members (i.e., a percentage or decimal) may be more precise in pinpointing the effects of team composition, the categorical assessment used in this study will lay the initial groundwork to examine team composition. Categorically conceptualizing this human-agent ratio also makes experimental manipulation possible by creating a small number of distinct, ecologically valid groups of varying team composition. A categorical conceptualization of this human-agent ratio also allows for inferences from human teams about how human team members process majority-minority characteristics in their team. Future efforts using a continuous conceptualization of this human-agent ratio are likely warranted, however composition in this study will be conceptualized as the ratio of agent team members to human team members. This approach classifies human-agent teams into three categories accordingly: balanced teams (i.e., equal numbers of human team members and agent team members), human majority teams (i.e., more human team members than agent team members), and agent majority teams (i.e., more agent team members than human team members).

*3.4 Team Composition and Trust Towards the Team*

Although human-agent teams have only recently gained traction, research on human behavior towards others who are different has had a long-standing history in the social sciences. To understand the level of trust a person has in their HAT as a whole, I draw on self-categorization theory (Tajfel, Bilig, Bundy, & Flament, 1971) to describe the categorization and identification processes someone perceives when placed in a group. Self-categorization theory explains that when people are placed into a social group, they have a tendency to classify all the members of the group based on observable characteristics (Tajfel, Bilig, Bundy, & Flament, 1971). This categorization process is most often done based on salient characteristics that differentiate the present individuals and places them into different groups. Subsequently, a person identifies which group they belong to and they form a preference to those with whom they share similarities (i.e., an in-group). Social identity theory (Tajfel & Turner, 1979) further explains that people tend to express favoritism towards members of their in-group such as increased levels of initial trust, even when they are personally unfamiliar with the individuals in their in-group (Navarro-Carillo, Valor-Segura, & Moya, 2018). Applied to a HAT, a human team member will interpret the human-agent dichotomy to create a human in-group and agent out-group. If trust towards a team is a result of the combined trust a person has with each member of the team (Costa & Anderson, 2011), then a HAT's composition affects a person's trust towards their overall team because of the varying number of in-group members present. This is likely because people tend to associate themselves with positive characteristics (Alicke & Gorovun, 2005), and by extension, associate those same positive characteristics to those who are similar to themselves (Rand & Wexley, 1975). Thus, if a team is composed of more similar individuals, then a person will associate more positive characteristics to their team members. In a HAT, if a person believes that humans are more trustworthy (Castro-González, Admoni, & Scassellati, 2016), and there are more human team members, then the person will trust the team more. Simply put, a human team member will trust their team more when there are more human team members compared to more agent team members.

*Hypothesis 1: (a) trust towards the team will be lower in agent majority teams than trust towards the team in balanced teams and human majority* *teams, (b) trust towards the team in balanced teams will be higher than trust towards the team in agent majority teams, but lower than trust towards the team in human majority teams, and (c) trust towards the team in human majority teams will be higher than trust towards agents in agent majority teams and balanced teams.*

*3.5 Team Composition and Trust Towards Agents*

Although the psychological processes that drive a person's trust towards their team (i.e., categorization, identification) are likely still in effect, changing the referent of trust from the team to the agent team members introduces an additional consideration that should be accounted for when attempting to predict a human team member's trust toward their agent team members. In particular, the interaction between two principal mechanisms, categorization processes and majority/minority dynamics, must be incorporated to understand the degree to which human team members trust their agent team members across different team composition categories.

First, regarding categorization processes, a person also classifies the members of a team who are not included in their ingroup into an out-group. In a HAT, arguably the most salient distinguishing characteristic between team members is that of human versus agent, and therefore a person is likely to observe the inherent human-agent dichotomy to form a human in-group and agent out-group. Although it has been discussed that people tend to express favoritism towards members of their in-group (Navarro-Carillo, Valor-Segura, & Moya, 2018), the corresponding opposite implication is that members of an out-group are often perceived with markedly lower levels of trust (Insko, Schopler, & Sedikides, 1998; Tajfel & Turner, 1979). Whereas one's trust towards their team is based on the collective trust resulting from the number of in-group members, trust towards agent team members is based on attitudes towards outgroup members.

Second, while an individual may classify themselves into a particular group to decide on their ingroup, they cannot control which group constitutes the majority of the team. Social identity theory posits that their perceptions of others may be driven by two processes related to this majority/minority identification: perceptions of threat and assessment of similar others (Tajfel & Wilkes, 1963). Regarding perceptions of threat, research has shown that when a person identifies as a member of the minority group, the detriments of negative out-group perceptions are amplified (Harstone & Augoustinos, 1995). This is often a result of the in-group perceiving an increased threat to the value of their input (Hornsey & Hogg, 2000). In other words, whenever majority/minority groups are perceived to exist, the extent to which an individual's negative perceptions toward outgroup members will increase is based on whether they identify as part of the majority group. However, this is not the only way that majority/minority identification influences a person's perceptions, as they also assess how similar they are to other members of a group. When a person identifies themselves as part of the minority group, they will perceive themselves as less similar to their team which underscores a "them" mentality (Tajfel & Billic, 1974). In other words, identifying as part of the minority group highlights the differences between an individual and their team members in such a way that they will perceive their outgroup team members in the majority to be a separate party that is less trustworthy.

Considering the interaction between outgroup categorization alongside majority/minority grouping bears important implications for the attitudes that a person has towards other groups present in a team. Depending on whether human team members (a person's ingroup) are the majority or minority, a person will differ in the degree to which they trust their agent team members (the out-group). For agent majority teams, human team members are likely to show lower levels of trust towards their agent team members (the majority outgroup) because the agents team members are their own separate group (low similarity) whose activities are a bigger input to the team (high threat). Although human team members in a human majority team may still hold negative perceptions of agent team members, a human team member will identify as part of the majority group and thus do not perceive the agent team members difference between themselves and their agent team members (high similarity) or perceive the input of their agent team member(s) to be as threatening to their contributions (low threat). As a result, individuals in a human majority team will likely show higher levels of trust towards their agent team members than individuals in an agent majority team. Lastly, whereas the previous team compositions are marked by majority/minority categorizations, balanced teams observe an equal number of human team members and agent team members. Although the human-agent dichotomy still exists, the effect of categorization is not exacerbated by majority/minority dynamics. In short, human team members in a balanced team will not perceive the intergroup context which filters changes in perception towards outgroup members and are likely to show higher levels of trust towards their agent team members than individuals in both agent majority teams and human majority teams.

*Hypothesis 2: (a) trust towards agents in agent majority teams will be lower than trust towards agents in balanced teams and human majority teams, (b) trust towards agents in human majority teams will be higher than trust towards agents in agent majority teams, but lower than trust towards agents in balanced teams, and (c) trust towards agents in balanced teams will be higher than trust towards agents in agent majority teams and human majority teams.*

*3.6. Agent Role and Trust Towards Agents*

Although previous empirical research has examined HATs in which the agent is the leader (e.g., RoboLeader; Chen & Barnes, 2014), few studies have compared how team processes, emergent states, and outcomes differ when the agent team member is the leader compared to when the agent team member is a subordinate. Prior research from the leadership literature suggests that a person's perceptions of their leader are tangled with their perceptions of other roles ascribed to them. In particular, research on women in leadership has often found that women are rated as less effective leaders (Grossman, Eckel, Komai, & Zhan, 2019). This can be explained by role congruity theory, which states that people are disliked more when they do not exhibit attributes that are socially characteristic of someone in their role, and this prejudice can be exacerbated when the characteristics of two roles are incompatible. (Eagly & Karau, 2002). For example, a woman leader fills the gender role of a woman (which is associated with helping and supportiveness) and the organizational role of a leader (which is associated with commanding and assertiveness) and is thus caught in a no-win situation as exhibiting characteristics of either side will be perceived as a failure to act within one of her roles.

Role congruity theory has a parallel implication for the role of agents in a HAT team. Stemming from their origins as a tool rather than teammate, the role of an agent is perceived to be supportive and supplementary to a human during a task (Lyons, Mahoney, Wynne, & Roebke, 2018). However, automation is continuing to technologically advance to enable agents to fill the role of a teammate or leader (Ososky, Schuster, Philips, & Jentsch, 2013). In situations where the agent is a leader, role congruity theory suggests that human team members will view an agent leader to be inappropriate because they are not serving as they were intended to as the team's supporting technological component. As research has shown that human team members often hold negative attitudes towards agent team members that fail to perform their role (Lyons et al., 2018), they are also likely to show lower levels of trust towards the agents when an agent is the leader.

*Hypothesis 3: Trust will be lower when an agent is the leader than when the agents are all subordinates.*

*3.7 Trust Towards Agents and Cooperative Behavioral Intention*

The next phase of the attitudinal process links an individual's formed attitudes to their plan of action. Prior research on trust in both traditional human relationships (Lewicki, Tomilson, & Gillespie, 2006; Mayer & Davis, 1995) and human-agent interactions (Hancock et al., 2011) have established the importance of trust in predicting a person's behavioral intention. In uncertain environments, trust plays an important role in the decisions a person makes (Park, Jenkins, & Jiang, 2008). The spirit of this tendency is also recurrent in the human-agent trust literature, as shown by empirical research (de Visser et al., 2016; Hancock et al., 2011) as well as recent theoretical proposals which posit that people act to varying levels of cooperation with their agents depending on their trust in the agents (de Visser, Pak, & Shaw, 2018; You & Roberts, 2018).

Given the importance of trust in predicting behavioral intent from both fields of literature, the present study assesses the relationship between an individual's trust and their behavioral intention to establish this link with the context of this study. In addition to corroborating prior research, assessing this relationship within this study's scenarios will provide evidence indicating whether a person's trust towards their agent team members does predict their cooperative behavioral intention exists.

*Hypothesis 4: An individual's trust towards agent team members will positively predict their cooperative behavioral intention.*

Piecing together the relationships above reflects the overall attitudinal process within the theory of planned behavior through which a HAT's composition and agent team members' roles ultimately affect an individual's intention to cooperate with their agent team members. Combining these theories and argumentation would indicate that trust is the primary vehicle that describes how HATs of varying compositions and agent roles influence the intentions of its human team members to cooperate with their agent team members. As the composition of a HAT and its agents' roles affect the level of trust an individual human team member will have towards their agent, and trust serves as the basis upon which an individual plan their decisions:

*Hypothesis 5: An individual's trust towards agent team members mediates the relationship between (a) team composition and the individual's cooperative behavioral intention as well as (b) agent roles and the individual's cooperative behavioral intention.*

# Chapter 4: Methods

This study used an experimental vignette methodology to manipulate a HAT's composition and the role of the agent team members in different written scenarios to understand how a human team member will plan to act based on their perceptions of the scenarios. Although vignette paradigms, especially those with a sentence-based format, are often criticized for their low external validity (Hughes & Huby, 2002), and seen as an artificial recreation of its true dynamic environment with low generalizability (Roehling, 1999), a well-designed written vignette can still provide data with a sufficient level of external validity. In addition to improving causal inference through a true, randomized experiment that is unfeasible for HAT field studies with pre-composed teams, vignettes hold a relevant and important role in understanding decision-making and judgment (Aguinis & Bradley, 2014; Rossi & Nock, 1982). Incorporating the recommendations from decision points provided by Aguinis & Bradley (2014) to guide the design of a vignette study (see Table 1), this study employed a between-subjects experimental design using a sentence-based vignette paradigm set in a futuristic military operation employing robot soldiers.

Table 1
*Vignette decision points from Aguinis & Bradley (2014)*

| Decision Point | Study Decision | Rationale |
|---|---|---|
| 1. Deciding whether EVM is a suitable approach | Yes | Manipulation, outcome is behavioral intention, circumstance limitations[a] |
| 2. Choosing the type of EVM | Paper people studies | Capture decision making processes with explicit processes and outcomes |
| 3. Choosing the type of research design | Between-person design | Study length |
| 4. Choosing the level of immersion | Futuristic scenario | Circumstantial limitations[a] |
| 5. Specifying the number and levels of the manipulated factors | 3 (Composition) x 2 (Role) | Theoretically driven variable decisions |
| 6. Choosing the number of vignettes | 6 | Combination of manipulated variables are orthogonal and not unrealistic |
| 7. Specifying the sample and number of participants | 226 MTurk Workers | A-priori power analyses, sample access |
| 8. Choosing the setting and timing for administration | Single session virtual survey | Standard survey procedure |
| 9. Choosing the best method for analyzing the data | ANOVA/Regression Frameworks | See proposed analyses |

*Note*. Aguinis & Bradley (2014) provide a 10$^{th}$ decision point related to presenting results which was not applicable during this proposal phase, but will be integrated in the final manuscript. [a]t the time of this study, the social distancing practices used to combat the COVID-19 pandemic limits laboratory access and technological ability.

### *4.1 Participants*

Based on a-priori power analyses of the hypotheses (see Table 2 for results of the full power analysis), 226 participants were recruited using CloudResearch. Some researchers have criticized the use of crowdsourcing platforms as a source of data, citing inter-related sample issues such as a small worker population, superworkers completing the majority of tasks, and the over-exposure of workers to the same measures (Stewart et al., 2015). In addition to these sample issues, recent concerns have grown about the use of repeated workers and automation polluting a study's data (Litman, 2018). However, in spite of the recent scares stemming from these various threats to the quality data collection, there is evidence that they are not as pervasive as many researchers believe they are (Sprouse, 2011; Tapped Out or Barely Tapped: Debunking Common Issues With MTurk's Participant Pool, n.d). There are also multiple methods of quality control to detect artificial survey data and insufficient effort responding which ameliorate these concerns. Accordingly, data cleansing procedures were used to filter out poor data (e.g., attention checks, manipulation checks, survey time cutoffs).

Table 2

*A-priori power analyses*

| Hypothesis | Analysis | Sample Size Needed |
|---|---|---|
| 1 | One-Way Analysis of Covariance (ANCOVA) | 175 |
| 2/3 | Two-Way Analysis of Covariance (ANCOVA) | 226 |
| 4 | Linear Regression | 46 |
| 5 | Mediation (Bootstrapping approach, PROCESS) | N/A |

*Note*. All power analyses are based on $\alpha = .05$ and effect size of 0.3 (moderate).

To be included in the study, participants had to meet three criteria. First, the participant must be a legally consenting adult (i.e., be 18+ years of age). Second, the participant must be located in the United States. Although the influence of culture and other international differences might be interesting for future studies, it is beyond the scope of this initial study. Lastly, the participant must have military experience. Because the setting of the vignette describes military scenarios, it is important that the participants possess military experience that they may draw on. Restricting the sample to individuals with military experience will also ensure that responses are rooted in the shared framework of military training, as opposed to the wide variety that would be observed from civilian responses.

The final sample consisted of 217 CloudResearch workers with military experience. Workers averaged 40.66 years of age (*SD* = 13.23), of which 72.4% identified as men (27.6% women, 0% non-binary or other) and 61.3% identified as White (17.5% Black, 7.4% Asian, 17.5% Latinx, 5.1% other). It is worth noting that this final sample falls 9 participants short of the proposed sample identified by a-priori power analyses (226) due to an error that is currently unspecified (though posited to be a cleaning or screening error resulting from Qualtrics' data exporting procedure). I will continue investigating and exploring these 9 lost data points at a later time, and reconduct all analyses accordingly.

*4.2 Procedure*

To simulate a HAT, written vignettes describing military scenarios were used to describe a team composed of both human team members and agent team members (see Appendix A for the vignettes). The military context was also chosen for its relevance to the current intentions of the U.S. army to integrate agent team members into its future operations (U.S. Army, 2020). Because modern technology has yet to fully develop robot soldiers and perfect their implementation, the vignettes situated the participant in a future military operation decades from now to enhance the believability of the scenario and the believability that technology has been improved to near perfect reliability. This futuristic context thus had the added purpose of controlling for perceptions of reliability, which research has shown to highly influence a person's trust (Fan et al., 2008).

As this study uses a between-persons design in order to study mutually exclusive conditions (i.e., a HAT can only have one of the three team composition levels; Atzmüller & Steiner, 2010), participants were randomly assigned to one of six manipulated conditions resulting from the combinations of the agent's role (leader or subordinate) and the team composition (agent majority, balanced, or human majority; see Table 3 for a synopsis of the experimental conditions).

Table 3
*Experimental Conditions*

| | Agent Majority | Balanced | Human Majority |
|---|---|---|---|
| Agent Leader | Agent is the leader Participant and 2 agents are subordinates | Agent is the leader Participant, 1 agent, and 1 human are subordinates | Agent is the leader Participant and 2 humans are subordinates |
| Agent Subordinate | Participant is the leader 3 agents are subordinates | Participant is the leader 2 Agents and 1 human are subordinates | Participant is the leader 1 agent and 2 humans are subordinates |

Apart from experimental manipulations, there were no differences in vignettes across conditions. Within each condition, participants read a vignette consisting of three scenarios with surveys placed between scenarios to measure the appropriate constructs. The first scenario explained the situation to the participant and provided context about the hypothetical combat mission and their hypothetical team members. Following this, the second scenario detailed the beginning of the mission and described a movement situation in which the team must work together to reach their destination. Finally, the third scenario described a hypothetical combat situation in which they confronted enemy soldiers in a firefight. By having separate scenarios for the various situations, the vignette formed a progressive, changing story that is more reflective of the scenarios it depicts to increase the participant's involvement (Pierce & Aguinis, 1997). Multiple scenarios which mirror the process of the natural experience they depict have also been shown to elicit more natural behaviors from the participant when compared to a single scenario (Hughes & Huby, 2002). Stringing the three vignettes together also allowed for an aggregation of behavioral intentions across multiple instances. In vignette studies, a participant's behavioral intentions are captured through the hypothetical decisions they make following a vignette (Rossi & Peters, 1982). Multiple vignettes thus allowed for behavioral intentions to be averaged across the different contexts, increasing its reliability. Although not a formal part of the hypotheses, the varying scenarios contained in the three vignettes also enable additional exploratory analyses to examine differences across the task contexts represented in each vignette.

*4.3 Measures*

Participants took four surveys throughout reading the vignettes. The first survey, taken after the context scenario, measured the participant's initial perceptions of the team prior to engaging in any fictional actions. The second and third surveys occurred after the movement and combat scenarios respectively. The study concluded with the fourth survey which measured individual differences such as demographics and relevant control variables identified from prior research. Although individual differences are typically captured at the beginning of a study, this study captured them at the conclusion of the experiment because measures of multiple agent-related attitudes and preconceptions are included which may prime the participant if the measures are presented before reading the vignettes. Details about the measures used in this study are described below (see Appendix B for the full measures).

**Trust (team)**. Trust towards the team was captured in the first three surveys using eight items from two subscales in Wildman and colleagues' (in development) trust measure. Items from this subscale were adapted to shift the referent to the team. Participants were instructed to rate the extent they have felt statements about their team using a 1 (*Not At All*) to 5 (*Very Much So*) scale. The trust in competence subscale contained 4 items which measure an individual's trust in their team's ability to perform, and included statements such as "Certain that your team will perform well?" and "Confident in your team's ability to complete a task?". The trust in intent subscale contained 4 items which measured an individual's trust in the team's social conscience, and included statements such as "Positive that your team will try and do what is best for everyone?" and "Convinced that you can rely on your team to try their hardest?". This measure demonstrated high internal reliability ($\alpha$ = .94).

Additionally, a single-item measure asked participants to indicate how much they trusted their team from 1 (*Distrust Very Much*) to 5 (*Trust Very Much*) will be used. This single-item measure will be used for exploratory analyses rather than analyzing the hypotheses, and was intended to be a direct method of measuring trust in a reflective approach (i.e., asking about trust itself; Coltman, Devinney, Midget, & Venaik, 2008).

**Trust (agents)**. An individual's trust towards their agent team members was captured in the first three surveys using both a validated scale and a sub-group measure. General trust towards the participant's agent team members was captured using Körber (2018)'s trust in automation scale. Because this study used a vignette methodology, items that required the participant to reflect on observed events were dropped (e.g., "I was able to understand why things happened", "The system state was always clear to me"). Three items remained and were adapted to fit the scenarios in the vignette, such as "I trust the [robot soldier(s)] [in this situation]". Participants were instructed to rate their agreement with these statements from 1

(*Strongly Disagree*) to 5 (*Strongly Agree*). This measure demonstrated high internal reliability ($\alpha = .90$).

In addition to the three-item trust measure adapted from Körber (2018), a sub-group measure of trust was used in the first three surveys. Because this study examines the differences in how a human team member interacts with other human team members compared to agent team members, this subgroup measure instructed participants to rate the degree to which they trust their leader (if applicable), the human team members, and agent team members from 1 (*Distrust Very Much*) to 5 (*Trust Very Much*). A true network-style rating of each team member separately was considered, however, because the study captures one person's perception of a scenario rather than the perceptions of all individuals in a team, a network of team member ratings cannot be captured. Furthermore, network measures assess as many of the individual relationships between dyads within a team as possible. Within the context of the vignette, participants would not be able to distinguish ratings between the individual members of their team (e.g., agent one versus agent two, human one versus human two). As such, individual ratings would not provide meaningfully interpretable data for a network analysis. The individual rating for the agent members of a team (i.e., a subordinate robot soldier or a robot soldier leader) were additionally used a single-item measure of trust towards agents for supplemental analysis.

**Cooperative behavioral intention**. In line with the purpose of an experimental vignette methodology, the participants' responses to the movement and combat scenarios within a vignette were used to measure their behavioral intention to cooperate with the agent(s) in their team. In each scenario, participants were asked questions about how they would interact with each of their team members in general and when they dissented from the participants opinion. Each question contained three behavioral choices which were designed to indicate an intent to counteract the target team member, ignore the target team member, or cooperate with the target team member. Cooperative behavioral intention was thus measured continuously using a three-point scale reflecting the degree of cooperative intent conceptualized in this study, and demonstrated low internal reliability ($\alpha$ = .55).

**Controls**. Prior research on human-agent interaction has found several proximal influences which have been known to affect a person's trust towards both other human team members and other agent team members (Nickerson & Riley, 2004; Schaefer, Hill, & Jentsch, 2018). Two relevant constructs, propensity to trust and attitudes toward artificial intelligence, were measured in the individual differences survey at the end of the study to avoid priming the participant's responses.

***Propensity to trust.*** Research from the trust literature has shown that some individuals are more likely to naturally trust other people to a higher degree (Mayer & Davis, 1999). This study accounts for this by measuring the participant's propensity to trust with Mayer & Davis's (1999) 8 item scale. Participants were instructed to rate their agreement with statements such as "One should be very cautious with strangers” and “Most people can be counted on to do what they say they will do” from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). This measure demonstrated low internal reliability ($\alpha$ = .64).

***Attitude towards artificial intelligence.*** In a similar vein, research has shown that individuals hold varying predispositions towards automation (Nickerson & Reilly, 2004). This study specifically used a four-item measure from Backonja and colleagues (2018) to capture the participant’s attitude towards artificial intelligence (A.I.). Although distinctions have been made between various terms for machine teammates within the literature (Chen & Barnes, 2014), one definition of A.I. (Russell & Norvig, 2016) aligns well with the major criteria for identifying an agent (i.e., technology, able to independently perform tasks, interacts with a task environment). Participants were presented with Russell & Norvig’s definition of A.I. in the instructions, then rate their agreement with statements such as “I would feel anxious if I was given a job or task where I had to use [AI].” using a 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) scale. This measure demonstrated moderate internal reliability ($\alpha$ = .76).

# Chapter 5: Results

To examine hypothesis 1a, 1b, and 1c, two ANCOVAs were run to determine the effect of team composition on trust towards the team after controlling for propensity to trust and attitudes towards A.I. Two separate ANCOVAs were conducted for each control variable due to the high correlation found between the control variables ($r = .62$, $p < .001$; see Table 4). First, controlling for propensity to trust, there was no significant difference in trust towards the team using the multi-item measure between the three types of team composition, $F(2, 213) = 2.44$, $p = .09$, partial $\eta^2 = .02$. Second, controlling for attitudes towards A.I., there was no significant difference in trust towards the team using the multi-item measure between the three types of team composition, $F(2, 213) = 2.82$, $p = .06$, partial $\eta^2 = .03$. Based on these results, hypotheses 1a, 1b, and 1c were not supported. The relationship between team composition and trust towards the team was also analyzed using the single-item measure of team trust. After controlling for propensity to trust, there were no significant differences in trust towards the team using the single-item measure between the three types of composition, $F(2, 213) = 2.75$, $p = .07$, partial $\eta^2 = .03$. After controlling for attitudes towards A.I., there was a significant difference in trust towards the team using the single-item measure between the three types of composition, $F(2, 213) = 3.27$, $p = .04$, partial $\eta^2 = .03$. Pairwise comparisons indicated that trust towards the team using the single-item measure was significantly greater for participants in human majority teams ($M = 4.30$) than participants in agent majority teams ($M = 3.98$). However, there was no

Table 4
*Descriptive Statistics and Correlations*

| Variable | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Trust towards Team (Scale) | 3.93 | 0.66 | — | | | | | | | | | |
| 2. Trust towards Team (Single) | 4.13 | 0.75 | .69* | — | | | | | | | | |
| 3. Trust towards Agents (Scale) | 3.80 | 0.76 | .67* | .66* | — | | | | | | | |
| 4. Trust towards Agents (Single) | 3.29 | 1.50 | .31* | .34* | .46* | — | | | | | | |
| 5. Cooperative Behavioral Intention | 2.14 | 0.52 | -.03 | -.01 | .07 | -.23* | — | | | | | |
| 6. Propensity to Trust | 3.61 | 0.60 | .16* | .22* | .42 | .11 | .14* | — | | | | |
| 7. Attitudes Towards A.I. | 3.59 | 0.72 | .02 | -.01 | .13 | -.06 | .03 | .62* | — | | | |
| 8. Military Experience | 8.03 | 6.69 | .07 | .02 | -.02 | .02 | -.05 | -.04 | .02 | — | | |
| 9. Automation Experience | 3.04 | 1.32 | -.08 | .16* | .17* | .06 | .13 | .53* | .37* | -.07 | — | |
| 10. Age | 40.66 | 0.79 | .17* | .09 | .05 | .10 | -.08 | -.13 | -.09 | .28* | -.35* | — |

* $p < .05$

significant difference in trust towards the team using the single-item measure between balanced teams (*M* = 4.12) and agent majority teams, nor balanced teams and human majority teams. Based on the single item measure of trust towards the team, hypothesis 1c was supported, but hypothesis 1a and 1b were not supported. To account for potential effects related to an individual's experience, further ANCOVAs were conducted to control for age, military experience, and automation experience using both the scale and single-item measure of trust towards the team. After controlling for the effects of military experience, there was a significant difference in trust towards the team using the single-item measure between the three types of composition $F(2, 214) = 3.11$, $p = .047$, partial $\eta^2 = .03$. Pairwise comparisons indicated that trust towards the team using the single-item measure was significantly greater for participants in human majority teams (*M* = 4.07) than participants in agent majority teams (*M* = 3.79). However, there was no significant difference in trust towards the team using the single-item measure between balanced teams (*M* = 3.95) and agent majority teams, nor balanced teams and human majority teams. Concerning age and automation experience however, no significant relationships were detected using either the scale or single-item measure of trust towards the team.

To examine hypotheses 2a-2c and hypothesis 3, two-way ANCOVAs were conducted to examine the effects of team composition and agent role on trust towards agents while controlling for attitudes towards A.I. and propensity to trust. First, after controlling for propensity to trust, there was a significant two-way interaction between team composition and agent role on trust towards agents, $F(2, 210) = 4.43$, $p = .01$, partial $\eta^2 = .04$. Analyses of simple main effects for team composition and agent role were then assessed to examine the effects of each independent variable at levels of the other independent variable. There was no significant effect of agent role on trust towards agents between human majority teams ($M = 3.85$), balanced teams ($M = 3.80$), or agent majority teams ($M = 3.72$). There was also no significant effect of team composition on trust towards agents when the agent was the leader ($M = 3.70$) compared to when the agent was a subordinate ($M = 3.88$). Although no significant simple main effects were found, it is worth noting that the differences between the six conditions trended in a similar direction as predicted in the hypotheses. Namely, trust towards agents was higher when the agents were subordinates for balanced and human majority teams, whereas trust towards agents was higher than an agent was the leader for agent majority teams (*see Figure 2*).

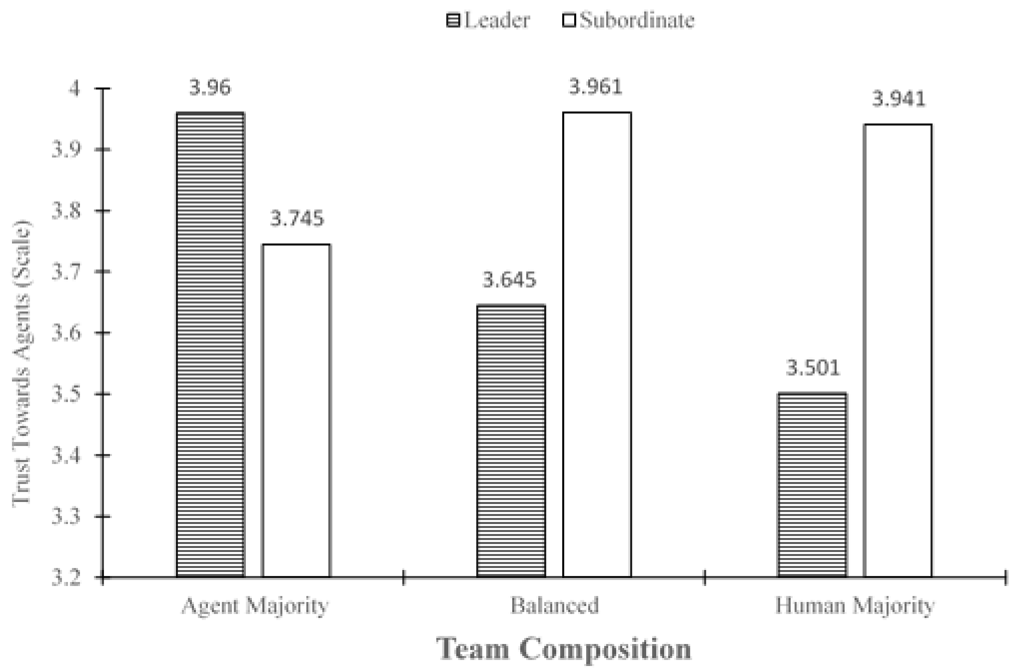


*Figure 2.* Interaction between team composition (agent majority, balanced, or human majority) and agent role (leader or subordinate) for trust towards agents (Körber, 2018) when controlling for propensity to trust.

Second, after controlling for attitudes towards A.I., there was a significant two-way interaction between team composition and agent role on trust towards agents, $F(2, 210) = 4.17$, $p = .02$, partial $\eta^2 = .04$. Thus, analyses of simple main effects for team composition and agent role were performed for each independent variable separately. There was no significant effect of agent role on trust towards agents in human majority teams ($M = 3.81$), balanced teams ($M = 3.83$), or agent majority teams ($M = 3.74$). There was also no significant effect of team composition on trust towards agents when the agent was the leader ($M = 3.70$) compared to when the agent was a subordinate ($M = 3.88$). Based on these results, hypotheses 2a, 2b, 2c, and 3 were unsupported. Similarly to the factorial ANCOVA controlling for propensity to trust, the differences between the six conditions trended in a similar direction to the hypotheses. Specifically, trust towards agents was higher when the agents were subordinates for balanced and human majority teams, whereas trust towards agents was higher than an agent was the leader for agent majority teams (see Figure 3). To account for potential effects related to an individual's experience, further ANCOVAs were conducted to control for age, military experience, and automation experience using both the scale and single-item measure of trust towards the team. Similarly to the results above, significant interaction effects were found for age ($F(2, 210) = 4.52$, $p = .01$, partial $\eta^2 = .04$), military experience ($F(2, 214) = 4.37$, $p = .01$, partial $\eta^2 = .04$) and automation experience ($F(2, 216) = 4.31$, $p = .02$, partial $\eta^2 = .04$), however no significant simple effects were found for any of these experience related control variables.

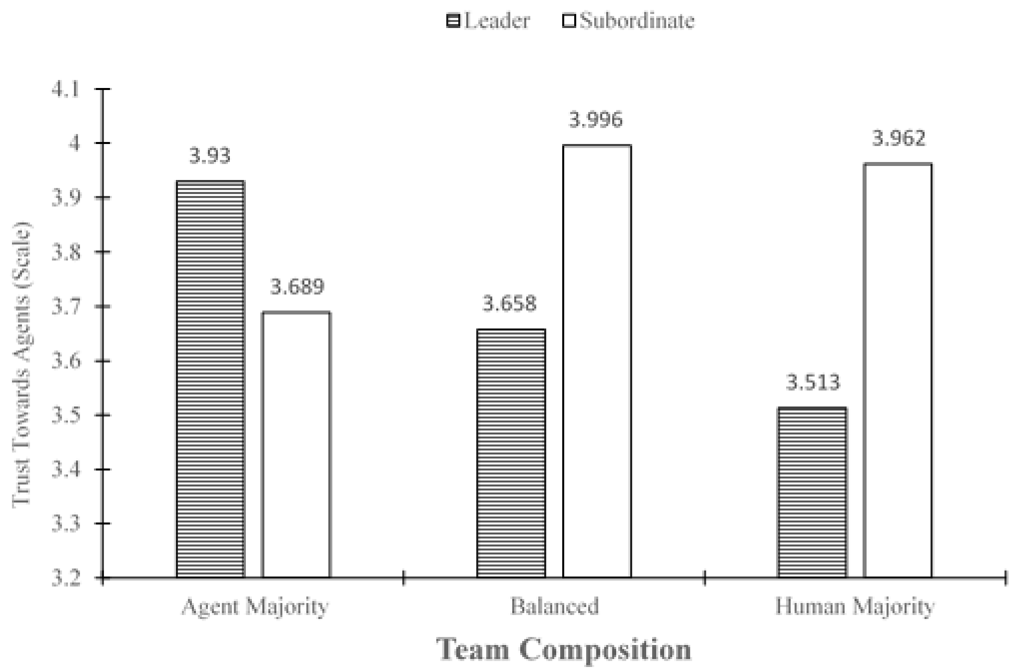


*Figure 3.* Interaction between team composition (agent majority, balanced, or human majority) and agent role (leader or subordinate) for trust towards agents (Körber, 2018) when controlling for attitudes towards A.I..

To examine hypothesis 4, a linear regression was conducted using trust towards agents to predict cooperative behavioral intention towards agents. The control variables, propensity to trust and attitudes towards A.I., were entered into the regression in the first step. The predictor, trust towards agents (scale), was entered into the second step. Results showed that trust towards agents did not significantly predict cooperative behavioral intention towards agents, $R^2 = .023$, $b = .005$, $F(2, 212) = 1.67$, $p = .18$. This regression was conducted again using the single-item measure of trust towards agents taken from the subgroup measure, however results did not show that the single-item measure of trust towards agents predicted cooperative behavioral intention towards agents, $R^2 = .023$, $b = .008$, $F(2, 212) = 1.67$, $p = .18$. To account for potential effects related to an individual's experience, additional linear regressions were conducted to control for age, military experience, and automation experience. However, no significant relationship was found between trust towards agents and cooperative behavioral intention towards agents for any of these regressions.

To examine hypothesis 5a and 5b, two mediation models were used to examine the indirect effects of team composition and agent role on cooperative behavioral intention through trust towards agents (see Table 4). First, regarding team composition, a mediation model was performed using a bootstrapping approach in the PROCESS macro for SPSS (Hayes, 2009). Unstandardized indirect effects were computed for 5,000 bootstrapped samples using a 95% confidence interval. Results showed that were no significant total ($p$ = .55) and direct ($p$ = .91) effects of team composition on cooperative behavioral intention. Additionally, there was no indirect effect of team composition on cooperation behavioral intention through trust towards agents as indicated by the 95% confidence interval (-.013 to .008), $b$ < -.001. Second, regarding agent role, the same mediation framework was performed using a bootstrapping approach in the PROCESS macro for SPSS. Unstandardized indirect effects were computed for 5,000 bootstrapped samples using a 95% confidence interval. Although results showed that both total ($p$ < .001) and direct ($p$ < .001) effects of team composition on cooperative behavioral intention were significant, evidence of a mediation model is inferred from indirect effects, which was not significant as indicated by the 95% confidence interval (-.001 to .057), $b$ = .016. Thus, neither hypotheses 5a nor 5b were supported.

*5.1 Exploratory Analyses*

The current study presented the opportunity to explore several additional analyses with potential to guide future research. First, the relationship between team composition and an individual's trust towards agents was examined within each scenario to explore whether additional task-based or environment-based factors affect a person's psychological processes. To examine the relationship between trust towards agents within each scenario while controlling for the effects of propensity to trust and attitudes towards A.I., six separate one-way ANCOVAs were conducted. Controlling for propensity to trust, there was no significant difference in trust towards agents within the context scenario ($F$ (2,212) = 1.40, $p$ = .25, partial $\eta^2$ = .01), within the movement scenario ($F$ (2,212) = .12, $p$ = 0.89, partial $\eta^2$ < .01), or the combat scenario ($F$ (2,212) = .02, $p$ = 0.98, partial $\eta^2$ < .01). Controlling for attitudes towards A.I., there was no significant difference in trust towards agents within the context scenario ($F$ (2,213) = .69, $p$ = 0.5, partial $\eta^2$ = .01), within the movement scenario ($F$ (2,213) = 0.44, $p$ = 0.64, partial $\eta^2$ < .01), or the combat scenario ($F$ (2,213) = 0.04, $p$ = 0.96, partial $\eta^2$ < .01). Based on these results, there is no evidence that the relationship between team composition and trust towards agents varies depending on the environmental context.

The relationship between trust towards agents and an individual's cooperative behavioral intention was also examined within the movement scenario and combat scenario using linear regression. This relationship could not be examined within the context scenario since participants did not respond to scenario-based questions measuring cooperative behavioral intention until after the movement scenario. A simple linear regression was calculated to predict an individual's cooperative behavioral intention based on their trust towards agents within the movement scenario, with propensity to trust and attitudes towards A.I. entered into the first step. Results indicated that trust towards agents did not predict cooperative behavioral intentions within the movement scenario, $R^2 = .045$, $b = .002$, $t(2, 214) = .03$ $p = .98$. The same regression framework was used to predict an individual's cooperative behavioral intention based on their trust towards agents within the combat scenario. Results indicated that trust towards agents did not predict cooperative behavioral intentions within the combat scenario, $R^2 < .01$, $b = -.01$, $t(2, 214) = -1.34$, $p = .89$. Based on these results, there is no evidence that the relationship between trust towards agents and cooperative behavioral intention varies depending on the environmental context.

Lastly, Wildman and colleagues' (in development) scale that was used to measure an individual's trust towards their team contains two sub-dimensions: trust in competence and trust in integrity. Future research may benefit from analyses comparing differences in specific sub-facets of team trust, and expand upon prior research findings that human team members often trust automation with technical tasks, but not decision-making tasks (Dzindolet et al., 2001). Four separate one-way ANCOVAs were conducted to examine the relationship between team composition and the two sub-dimensions of team trust while accounting for two control variables. When controlling for propensity to trust, there was no significant difference between the three types of team composition for an individual's trust in their team's competence ($F$ (2, 213) = 2.24, $p$ = .11, partial $\eta^2$ = .02) or integrity ($F$ (2, 213) = 2.24, $p$ = .11, partial $\eta^2$ = .02). When controlling for attitudes towards A.I., there was no significant difference between the three types of team composition for an individual's trust in their team's competence ( $F$ (2, 213) = 2.56, $p$ = .08, partial $\eta^2$ = .02) or integrity ($F$ (2, 213) = 2.61 $p$ = .08, partial $\eta^2$ = .02). Based on these results, team composition does not differentially predict an individual's trust in their team's competence compared to their trust in the team's integrity.

# Chapter 6: Discussion

The present study found few pieces of conclusive evidence regarding the effect of team composition or agent roles on the attitudes and behavioral intentions that human team members hold towards their agent team members in HATs. Although a few significant findings were found from hypothesis and exploratory analyses, several theoretical and methodological limitations of the current study may be culprits that hindered the detection of significant relationships. Several of these limitations bear particular relevance to certain hypotheses as they pertain to its focal variables, and are respectively interpreted and discussed for each hypothesis below.

Hypothesis 1 posited that trust towards the team differed between the three types of team composition, and was partially supported. Although no significant differences were found between balanced teams and other team composition types, individuals in human majority teams trusted the team significantly more than individuals in agent majority teams. This difference may indeed suggest that human team members perceive their team differently depending on the whether the majority of the team is composed of human team members or agent team members. Theoretically, the categorical approach taken to conceptualize team composition in this study may have been too simplistic. Although the essence of the three categories of team composition used in this study was a good foothold for beginning research into team composition, it may not have been comprehensive enough to capture the qualities that would make team composition a relevant predictor of trust and behavioral intention. Social identity theory (Tajfel & Turner, 1979) and research on group dynamics (Hornsey & Hogg, 2000; Insko, Schopler, & Sedikides, 1998; Navarro-Carillo, Valor-Segura, & Moya, 2018) pinpoint the importance of identifying with an in-groups versus out-group, which is captured with the categorical approach to composition used in this study, the rationale for hypothesis 1 depended on a continuous representation of the number of in-group members in a group. Specifically, the theoretical reason underlying the predictions made in hypothesis 1 was formed on the basis that the bigger an individual's in-group was in a group, the more they would trust the group as a whole. Thus, this hypothesis may be more suitably tested using a continuous form of team composition with varying in-group sizes (i.e. differing numbers of human team members). However, it is worth noting that the trust scores still trended in the predicted direction and were marginally significant when using the trust towards team scale ($p$ = .06). This may suggest issues with the study's power, and it may be that this relationship would have been significant with a higher sample size.

Hypothesis 2 posited that an individual's trust towards agents differed between the three types of team composition. Although this prediction was not supported, theoretical and methodological limitations of the current study may explain why evidence of this relationship could not be detected. In terms of theoretical limitations, the rationale linking team composition and trust towards agents depended on outgroup perception. More importantly, in addition to actually identifying members of one's in-group, an individual must actually process and detect the existence of an outgroup. The argumentation for the relationship between team composition and trust towards agents is based on theory describing an “us” versus “them” mentality as explained by group identification and emphasized by majority/minority dynamics. If this polarity is not perceived, then these underlying reasons which might drive differences in trust towards agents may not occur. Although this in-group/out-group divide should be seemingly obvious in a HAT, the methodological limitations of a vignette study may hinder this. Whereas human team members in a real-world HAT must actually interact with agent team members through their technological medium, and thus register noticeable differences between themselves and their agent team member, participants in this vignette study must fictitiously envision the robot soldiers in this team. Aside from stating that there are robot soldiers in this hypothetical mixed team, there are no additional cues which would hone the participant into processing the human-agent divide in the team. It is conceivable that a participant might not psychologically register in-group/out-group categories because of the ambiguous nature of written vignettes which leave room for interpretation, at which point the theoretical underpinning of hypothesis 2 would not hold. Additionally, there may be multiple measurement issues in regards to capturing trust towards agents (which are discussed in further depth below in the limitations section). Although the interaction effects found in this analysis were not hypothesized, it provides a fascinating opportunity for future research to understand why the relationship between agent role and team composition differed for agent majority teams.

Hypothesis 3 predicted that an individual's trust towards agents would be lower when an agent was the leader compared to when the agents were all subordinates. The results of this study did not support hypothesis 3. In addition to the measurement issues in capturing trust towards agents, one alternative explanation as to why a relationship between agent roles and trust towards agents was not found may be due to the confounds of team composition. A significant interaction effect was found between team composition and agent role on trust towards agents in the factorial ANCOVA conducted, but no simple main effects or main effects were found. Balanced and human majority teams had other human team member subordinates, and it is unclear what effect this might have on the perceptions of participants who were the team leader. It is possible that participant leaders in these scenarios were influenced by the presence of other human team members, however it is unclear how this influence might manifest. Thus, in addition to issues with measuring trust towards agents, team composition may confound the effect of agent roles due to added human subordinates in the balanced and human majority conditions.

Based on prior research, hypothesis 4 posited that an individual's trust towards agents predicted their cooperative behavioral intention towards agents. Given the established nature of this relationship from the literature on trust in HATs, it was surprising that this hypothesis was unsupported. However, this may lend further support to the measurement concerns regarding trust towards agents and cooperative behavioral intention.

To address the teaming problem in HATs, hypothesis 5 posited that the relationship between team composition and cooperative behavioral intention, as well as agent role and cooperative behavioral intention, were both explained by an individual's trust towards agents. This mediation model was not supported. In the same vein, it is likely that these indirect effects could not be detected for the reasons listed for the above hypotheses, just as it is likely that none of the exploratory analyses yielded significant findings due to these same constraints.

*6.1 Limitations*

In addition to the specific limitations pertaining to the hypotheses, several theoretical, methodological, and sample limitations of the study as a whole may further explain the lack of significant findings. At large, the study is grounded in the theory of planned behavior (Azjen, 1985), which posits that attitudes, perceived control, and subjective norms are the three factors that predict an individual's behavioral intention. Of these three factors, this study solely examines the attitudinal component of the theory by focusing on how trust predicts cooperative behavioral intention. Although the hypotheses in this study did not provide evidence that team composition affects cooperative behavioral intention through trust, Azjen (1985) states that the components of the theory are not necessarily isolated predictors of behavioral intention, and thus perceived control and subjective norms may also influence an individual's trust. For example, an organization may provide cues about the trustworthiness of the agent through formal (e.g., training) or informal (e.g., socialized beliefs) means. Accounting for these components of the theory may improve the ability to isolate the relationship between team composition, trust, and cooperative behavioral intention, as well as offer alternative mediation models using perceived control and subjective norms to elucidate a relationship between team composition and cooperative behavioral intention. Another theoretical issue that likely requires further investigation concerns the time that it takes for the relationships in the study to form. Research on the development of trust in automation states that trust is adjusted after being exposed to a stimulus, and thus trust is temporally calibrated (Schaeffer, Hill, & Jentsch, 2018). Although it is unclear how long it takes for trust to adjust based on the stimulus in question, it may be that the cross-sectional nature of this study does not capture trust at the appropriate time to measure trust as it is calibrated to the stimulus in question (e.g., the scenarios).

Methodological limitations present multiple important issues that likely limited the ability to detect a relationship. A prominent methodological issue relates to the inherent limitations of a written vignette methodology for HAT research. To date, no written, experimental vignette study of HATs exists, as experimental research frequently opts to test laboratory trials involving a real agent that a participant interacts with. Subsequently, findings from the literature are based on data in which human team members in a HAT interact with their agent team to actually conduct a task. Although written vignette methodologies provide for highly controlled experiments through careful and deliberate manipulations of independent variables in question, it is likely that participants engaging in an imaginary scenario are missing out on experiences that require actual interaction with an agent team member. Without actually interacting with an agent, a human team member may be limited in their ability to fully infer and develop attitudes towards the agents, and subsequently act on these formed attitudes. Aguinis and Bradley (2014) state that it is important that participants are sufficiently immersed within a vignette to react as they would in actual scenarios that the vignettes mimic. This criteria is difficult to meet, as it is challenging to provide a sufficient level of immersion for HATs using a written vignette methodology. It may also be the case that the "perfect reliability" controlled for across all vignettes is different from the high levels of reliability which we observe in current tech. Prior experimental research has made inferences based on data using real, current technology which is imperfect. Although this perfect reliability controls for the influence of reliability on the relationships in this study, it is possible that introducing perfect reliability to HAT research operates on a different paradigm than current research as it removes perceptions of variability in performance that trust in automation research has heavily focused on (deVisser, Pak, & Shaw, 2018).

The lack of actual agent interaction in a vignette study provides additional measurement issues. First, regarding the measurement of trust towards agents, the original scale developed and validated by Körber (2018) was significantly adapted. Specifically, fifteen items were dropped from the original scale as they required the participant to draw on experiences with the agent (e.g. "I was always able to understand why things happened"). Although the scale demonstrated high reliability ($\alpha$ = .90), the shortened scale has not been validated and it is uncertain how many of the dropped items were core to the purpose of the scale. This also prompts questions as to whether reading a statement about an agent's action is sufficiently comparable to observing the agent's action itself. For example, can an individual reading a sentence stating their agent team member is failing to win its fights comparably process the same information as an individual watching their agent team member lose in combat? Based on prior research into trust and reliability which manipulate the details of these failures to see how human team members react, there may be more nuance to an instance of failure than just identifying a state of success versus failure (deVisser, Pak, & Shaw, 2018). Lastly, trust towards agents may be a new construct that is distinct from trust towards automation. As the premise of the agents in a human-agent team is that agents act as team members, capturing trust towards agent team members may require measures which reflect this mindset and move beyond items about reliability, and towards items similar to human-human trust (i.e., relationship-focused).

The reliability and validity of the cooperative behavioral intention measures are also suspect, as it is captured through the simulated response options provided to the participant for the context of this vignette. The items indicating cooperative behavioral intention demonstrated low reliability ($\alpha$ = .55). Although the three response options provided consistently represented the same range of cooperative behaviors for each behavioral intention item (e.g., counteracting, ignoring, and cooperating), the validity of these items as a composite indication of cooperative behavioral intention warrants further evaluation. Specifically, it is worth considering whether or not these behaviors actually map onto a continuous spectrum. It may be the case that each of these various behaviors should be viewed as an individual outcome, and thus should each be individually predicted. Additionally, because cooperative behavioral intention is inferred on a scale of one to three, there is a limit to the variance in cooperative intent that can be expressed. A participant thus cannot express different levels of counteracting behavior and cooperative behavior, as they are limited to simply indicating counteractive intent (coded as one) or cooperative intent (coded as three) without nuance at the extreme ends. Additionally, cooperative behavioral intent was inferred by averaging all items within a vignette (i.e. across all scenarios), however it may be the case that these observations should not be equally weighted. For example, it is unclear whether choosing to deter from a robot's soldier's suggested path is equally representative of low cooperative intention as choosing to overriding a robot soldier during a fight. Accurately capturing cooperative behavioral intention may require deeper investigation into what cooperation really means beyond these three points of alignment (e.g. counteracting, ignoring, cooperating).

Lastly, two sampling limitations which may have affected analyses are worth noting. First, variances between the six conditions in this study were not equal for trust towards agents, which fails the assumption of homogeneity of variance required of ANCOVA analyses and affects hypotheses 2 and 3. Second, the sampling population was broad given the specificity of the scenario. The vignettes describe an infantry combat scenario, however the sample was taken from any individuals with military experience from any branch. It would have been ideal to sample from individuals with boots-on-the-ground experience, however the sample pool for this would have been small and implausible to obtain a sufficient sample size within a reasonable time.

*6.2 Future Research*

Although the present study did not provide evidence regarding the effect of team composition, it should not be interpreted as a deterrent from future research on team composition in HATs. Research on team composition may be viable in addressing the teaming problem in HATs. Within the vein of this study, multiple limitations were identified and thus multiple suggestions for future research arise. First, and perhaps most prominently, this study may benefit from in-person laboratory experiments for the theoretical and methodological reasons discussed above. Given the nature of HATs, and how the HAT literature and its measures are built upon laboratory trials in which participants interact with the technological embodiment of the agent in some capacity (Hancock et al., 2011), accurately capturing the psychological processes that occur during human-agent interaction likely requires physical and psychological interaction. A future iteration of this study may also benefit from an in-depth examination of open-ended responses to the type of hypothetical scenarios used in this study. Using qualitative data can be insightful for identifying a trend or phenomenon, or understanding why certain events may be unfolding the way that they do (Briner et al., 2011). By exploring themes in open-ended responses to these hypothetical scenarios, future researchers may be able to examine how behavioral intentions are formed and how they are connected. Doing so may hone future research on predicting cooperation more appropriately as it reflects real-world behavioral instances. Similarly, future research based on this study's framework to address the teaming problem would benefit from testing the components of this study's behavioral intention scale individually. Returning to literature on automation usage (Parasuraman & Riley, 1997), which categorically identifies four types of usage behavior, it may be more elucidating to treat each type of behavior as its own independent outcome. Rather than trying to predict a range of cooperation as inferred by the participants choice, separately predicting the occurrence of the specific behaviors (e.g., cooperation, ignoring, counteracting) may better align with prior research and provide specific actionable implications for addressing the teaming problem in HATs as well. It may also be worthwhile for future researchers to parse out high cooperation as well based on the appropriateness of the cooperation in line with the distinction between use and misuse (Meyer & Lee, 2013).

Future research broadly interested in team composition in HATs could also benefit from independently examining each of the types of team composition in this study. Rather than comparing agent majority, balanced, and human majority teams, individually the mechanisms within each team would allow for a more elaborate analysis of how an individual processes the team's context and reacts to agent team members (i.e., identifying mediators that are unique to agent majority teams that explain how trust is calibrated). Doing so would also enable a continuous examination of each type of team composition on its own, and allow researchers to explore how or if the specific type of team composition changes with varying size (i.e., comparing agent majority teams with two agents versus 5 agents).

Alternatively, if research were to continue using the categorical framework of team composition that was conceptualized in this study, the interaction effect found in the factorial ANCOVA that tested hypotheses 2 and 3 provides a curious opportunity for understanding the nuance between agent role and team composition when it comes to predicting a human team member's trust towards agents. If this interaction effect could be replicated, it would provide both practical and theoretical information in regard to how certain configurations of a team's inputs (e.g., agent role and team composition) will influence the trust that human team members will have towards agents. Future research could study this "humans leading humans, agents leading agents" phenomenon with theoretically grounded rationale to provide meaningful insight into understanding how team composition and agent roles influence the human team members in a HAT.

Lastly, this study was driven by the theory of planned behavior yet only examined the attitudinal piece of the theory to leverage the extensive trust research that is established within the literature. Future research interested in addressing the teaming problem in HATs should consider the other components of the theory of planned behavior, as it is possible that the subjective norms and perceived control regarding agent team member interaction would influence the attitudes the individual has towards the agents and towards engaging in specific behaviors with the agents. More elaborate models which examine the interplay between these three components of the theory of planned behavior, as well as test formal hypotheses regarding each individual component, may be needed to uncover the effects of any contextual influences of a HAT on a human team members behaviors.

*6.3 Conclusion*

The present study attempted to address the teaming problem in HATs by examining the relationship between team composition, agent roles, trust, and cooperative behavioral intention as explained through the theory of planned behavior. Although this study did not find any evidence supporting its hypotheses, multiple limitations and recommendations for future research should not discourage future attempts to study team composition in HATs.

References


Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351-371.

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. *The Self in Social Judgment*, *1*, 85-106.

Anjomshoae, S., Najjar, A., Calvaresi, D., & Främling, K. (2019). Explainable agents and robots: Results from a systematic literature review. In *18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019* (pp. 1078-1088). International Foundation for Autonomous Agents and Multiagent Systems.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11-39). Springer, Berlin, Heidelberg.

Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology*.

Backonja, U., Hall, A. K., Painter, I., Kneale, L., Lazar, A., Cakmak, M., ... & Demiris, G. (2018). Comfort and Attitudes Towards Robots Among Young, Middle-Aged, and Older Adults: A Cross-Sectional Study. *Journal of Nursing Scholarship*, *50*(6), 623-633.

Battiste, V., Lachter, J., Brandt, S., Alvarez, A., Strybel, T. Z., & Vu, K. P. L. (2018, July). Human-Automation teaming: Lessons learned and future directions. In *International Conference on Human Interface and the Management of Information* (pp. 479-493). Springer, Cham.

Bell, S. T., Brown, S. G., Colaneri, A., & Outland, N. (2018). Team composition and the ABCs of teamwork. *American Psychologist*, *73*(4), 349.

Billings, D. R., Schaefer, K. E., Hancock, P. A., Kocsis, V., Barrerra, M., Cook, J., & Ferrer, M. (2012a). *Human Animal Trust as an Analog for Human-Robot Trust: A review of current evidence* (Report No. ARL-TR 5949). Aberdeen Proving Ground, Maryland: Army Research Laboratory.

Bradshaw, J. M., Feltovich, P. J., Johnson, M. J., Bunch, L., Breedy, M. R., Eskridge, T., ... & Uszok, A. (2008, May). Coordination in human-agent-robot teamwork. In *2008 International Symposium on Collaborative Technologies and Systems* (pp. 467-476). IEEE.

Burke, J. L., & Murphy, R. R. (2004, September). Human-robot interaction in USAR technical search: Two heads are better than one. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)* (pp. 307-312). IEEE.

Cahya, D. E., & Giuliani, M. (2018). Towards a cognitive architecture incorporating human feedback for interactive collaborative robots. *Lecture Notes in Computer Science*, *10965*.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Current Issues in Individual and Group Decision-Making* (pp. 221-246). Hillsdale, NJ: Erlbaum.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. *Team Effectiveness and Decision Making in Organizations*, *333*, 380.

Casper, J., & Murphy, R. R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *33*(3), 367-385.

Castro-González, Á., Admoni, H., & Scassellati, B. (2016). Effects of form and motion on judgments of social robots′ animacy, likability, trustworthiness and unpleasantness. *International Journal of Human-Computer Studies*, *90*, 27-38.

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.

Chen, J. Y., & Terrence, P. I. (2009). Effects of Imperfect Automation and Individual Differences on Concurrent Performance of Military and Robotics Tasks in a Simulated Multitasking Environment. *Ergonomics*, *52(8)*, 907–920

Chen, J. Y., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical issues in ergonomics science*, *19*(3), 259-282.

Chakraborti, T., Kambhampati, S., Scheutz, M., & Zhang, Y. (2017). Ai challenges in human-robot cognitive teaming. arXiv:1707.04775.

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.

Chiou, E. K., & Lee, J. D. (2016). Cooperation in human-agent systems to support resilience: A microworld experiment. *Human factors*, *58*(6), 846-863.

Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. *Advances in Human Performance and Cognitive Engineering Research*, *2*, 1-12.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250-1262.

Costa, A. C., & Anderson, N. (2011). Measuring trust in teams: Development and validation of a multifaceted measure of formative and reflective indicators of team trust. *European Journal of Work and Organizational Psychology*, *20*(1), 119-154.

Czaja, S. J., & Sharit, J. (1998). Age differences in attitudes toward computers. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 53(5)*, 329-340.

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied, 22(3)*, 331.

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation'to 'autonomy': the importance of trust repair in human–machine interaction. *Ergonomics*, *61*(10), 1409-1427.

de Visser, E., Parasuraman, R., Freedy, A., Freedy, E., & Weltman, G. (2006, October). A comprehensive methodology for assessing human-robot team performance for use in training and simulation. In *Proceedings of The Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 25, pp. 2639-2643). Sage CA: Los Angeles, CA: SAGE Publications.

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*, 719–735.

Demir, M., Amazeen, P. G., McNeese, N. J., Likens, A., & Cooke, N. J. (2017, September). Team coordination dynamics in human-autonomy teaming. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 236-236). Sage CA: Los Angeles, CA: SAGE Publications.

Demir, M., Cooke, N. J., & Amazeen, P. G. (2018). A conceptual model of team dynamical behaviors and performance in human-autonomy teaming. *Cognitive Systems Research, 52*, 497-507.

Demir, M., McNeese, N. J., & Cooke, N. J. (2016, March). Team communication behaviors of the human-automation teaming. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (pp. 28-34). IEEE.

Demir, M., McNeese, N. J., Cooke, N. J., Ball, J. T., Myers, C., & Frieman, M. (2015, September). Synthetic teammate communication and coordination with humans. In *Proceedings of The Human Factors And Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 951-955). Sage CA: Los Angeles, CA: SAGE Publications.

Dunne, B. J. (1998). Gender differences in human/machine anomalies. *Journal of Scientific Exploration, 12(1),* 3-55.

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), 147-164.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573.

Endsley, M. R. (2015). Autonomous Horizons: System Autonomy in the Air Force- A Path to the Future. *United States Air Force Office of the Chief Scientist, AF/ST TR*, 15-01.

Evers, V., Maldonado, H., Brodecki, T., & Hinds, P. (2008, March). Relational vs. group self-construal: untangling the role of national culture in HRI. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 255-262). IEEE.

Eyssel, F., & Hegel, F. (2012). (s) he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology*, *42*(9), 2213-2230.

Fan, X., & Yen, J. (2010). Modeling cognitive loads for evolving shared mental models in human–agent collaboration. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 41*(2), 354-367.

Fan, X., Oh, S., McNeese, M., Yen, J., Cuevas, H., Strater, L., & Endsley, M. R. (2008, January). The influence of agent reliability on trust in human-agent collaboration. In *Proceedings of the 15th European conference on Cognitive ergonomics: the ergonomics of cool interaction* (p. 7). ACM.

Fan, X., Sun, B., Sun, S., McNeese, M., & Yen, J. (2006, May). RPD-enabled agents teaming with humans for multi-context decision making. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems* (pp. 34-41). ACM.

Fiore, S. M., & Wiltshire, T. J. (2016). Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in Psychology, 7*, 1531.

Fong, T., & Nourbakhsh, I. (2005). Interaction challenges in human-robot space exploration.

Gabler, V., Stahl, T., Huber, G., Oguz, O., & Wollherr, D. (2017, May). A game-theoretic approach for adaptive action selection in close proximity human-robot-collaboration. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 2897-2903). IEEE.

Gao, F. (2013, February). Modeling teamwork of multi-human multi-agent teams. In *Proceedings of the 2013 conference on Computer supported cooperative work companion* (pp. 47-50). ACM.

Gombolay, M. C., Huang, C., & Shah, J. (2015, September). Coordination of human-robot teaming with human task preferences. In 2015 *AAAI Fall Symposium Series*.

Gombolay, M. C., Gutierrez, R. A., Clarke, S. G., Sturla, G. F., & Shah, J. A. (2015). Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots*, *39*(3), 293-312.

Goodrich, M. A., & Schultz, A. C. (2008). *Human-robot interaction: a survey*. Now Publishers Inc.

Goodrich, M. A., & Yi, D. (2013, July). Toward task-based mental models of human-robot teaming: A bayesian approach. In *International conference on virtual, augmented and mixed reality* (pp. 267-276). Springer, Berlin, Heidelberg.

Gouman, R., Kempen, M., Wijngaards, N., French, S., Tomaszewski, B., & Zobel, C. (2010, May). Actor-agent team experimentation in the context of incident management. In *Proc. 7th Intl. Conf. on Information Systems for Crisis Response and Management, ISCRAM*.

G. R. Semin, "Grounding communication: Synchrony," in S*ocial psychology: Handbook of basic principles, 2nd ed*, A. W. Kruglanski and E. T. Higgins, Eds. New York, NY, US: Guilford Press, 2007, pp. 630–649

Grossman, P. J., Eckel, C., Komai, M., & Zhan, W. (2019). It pays to be a man: Rewards for leaders in a coordination game. *Journal of Economic Behavior & Organization*, *161*, 197-215.

Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In *Advances in Experimental Social Psychology* (Vol. 8, pp. 45-99). Academic Press.

Hancock, P. A., Billings, D. R., Oleson, K. E., Chen, J. Y., De Visser, E., & Parasuraman, R. (2011). *A meta-analysis of factors influencing the development of human-robot trust*. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE.

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, teams, and task performance: Changing effects of surface-and deep-level diversity on group functioning. *Academy of Management Journal*, *45*(5), 1029-1045.

Hartstone, M., & Augoustinos, M. (1995). The minimal group paradigm: Categorization into two versus three groups. *European Journal of Social Psychology*, *25*(2), 179-193.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*(3), 451-470.

Hobbs, A. (2008). An overview of human factors in aviation maintenance. *ATSB Safty Report, Aviation Research and Analysis Report AR*, *55*, 2008.

Hoffman, R. R., J. D. Lee, D. D. Woods, N. Shadbolt, J. Miller, and J. M. Bradshaw (2009). "The Dynamics of Trust in Cyberdomains." *IEEE Intelligent System, 24(6)*, 5–11.

Hornsey, M. J., & Hogg, M. A. (2000). Assimilation and diversity: An integrative model of subgroup relations. *Personality and Social Psychology Review*, *4*(2), 143-156.

Hughes, R., & Huby, M. (2002). The application of vignettes in social and nursing research. *Journal of Advanced Nursing, 37*, 382-386.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Reviews of Psychology*, *56*, 517-543.

Insko, C. A., Schopler, J., & Sedikides, C. (1998). Differential distrust of groups and individuals. *Intergroup cognition and intergroup behavior*, 75-107.

Jehn, K. A., & Bezrukova, K. (2010). The faultline activation process and the effects of activated faultlines on coalition formation, conflict, and group outcomes. *Organizational Behavior and Human Decision Processes*, *112*(1), 24-42.

Jennings, N. R., Moreau, L., Nicholson, D., Ramchurn, S. D., Roberts, S., Rodden, T., & Rogers, A. (2014). Human-agent collectives. *Communications of the ACM, 57(12)*, 80-88.

Jentsch, F. (2016). *Human-Robot Interactions in Future Military Operations*. CRC Press.

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.

Kaminka, G. A., Pynadath, D. V., & Tambe, M. (2002). Monitoring teams by overhearing: A multi-agent plan-recognition approach. *Journal of Artificial Intelligence Research*, *17*, 83-135.

Keen, P. G. (1980). Decision support systems: Translating analytic techniques into useful tools. *Sloan Management Review (pre-1986)*, *21*(3), 33.

Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. Routledge.

Körber, M. (2018, August). Theoretical considerations and development of a questionnaire to measure trust in automation. In *Congress of the International Ergonomics Association* (pp. 13-30). Springer, Cham.

Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In WC Borman, DR Ilgen, & RJ Klimoski (Eds.), Handbook of psychology: Industrial and organizational psychology (Vol. 12, 333-375).

Kozlowski, S. W. J., & Klein, K. J. (2000). *A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes.* In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (p. 3–90). Jossey-Bass.

Krippendorff*, K. (2004). Intrinsic motivation and human-centred design. *Theoretical Issues in Ergonomics Science*, *5*(1), 43-72.

Kumar, K., & Thakur, G. S. M. (2012). Advanced applications of neural networks and artificial intelligence: A review. *IJ Information Technology and Computer Science*, *6*, 57-68.

Lau, D. C., & Murnighan, J. K. (2005). Interactions within groups and subgroups: The effects of demographic faultlines. *Academy of Management Journal*, *48*(4), 645-659.

Lee, J. D. (2006). Human factors and ergonomics in automation design. In *G. Salvendy (Ed.), Handbook of Human Factors and Ergonomics* (pp. 1570–1596). Hoboken, NJ: Wiley.

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

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*(1), 153-184.

Leng, J., Li, J., & Jain, L. C. (2008, September). A role-based framework for multi-agent teaming. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 642-649). Springer, Berlin, Heidelberg.

Levine, J. M., & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology*, *41*(1), 585-634.

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

Litman, L. (2018). Concerns about bots on Mechanical Turk: Problems and solutions. Retrieved May 23, 2020.

Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 1 - 10.

Lyons, J. B., & Guznov, S. Y. (2019). Individual differences in human–machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science, 20(4)*, 440-458.

Lyons, J. B., Mahoney, S., Wynne, K. T., & Roebke, M. A. (2018, March). Viewing machines as teammates: A qualitative study. In *2018 AAAI Spring Symposium Series*.

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.

Marinaccio, K., Kohn, S., Parasuraman, R., & De Visser, E. J. (2015, June). A framework for rebuilding trust in social automation across health-care domains. In *Proceedings of the international symposium on human factors and ergonomics in health care* (Vol. 4, No. 1, pp. 201-205). Sage India: New Delhi, India: SAGE Publications.

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), 356-376.

Marquez, J. J., Chang, M. L., Beard, B. L., Kim, Y. K., & Karasinski, J. A. (2018). Human-Automation Allocations for Current Robotic Space Operations. NASA Technical Reports Server.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, *84*(1), 123–136.

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Englewood Cliffs, NJ: Prentice-Hall.

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors, 60*(2), 262-273.

Mendonça, D., Brooks, J. D., & Grabowski, M. (2014). Linking team composition to team performance: an application to postdisaster debris removal operations. *IEEE Transactions on Human-Machine Systems*, *44*(3), 315-325.

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.

Meszaros, E. L., Le Vie, L. R., & Allen, B. D. (2018). Trusted Communication: Utilizing Speech Communication to Enhance Human-Machine Teaming Success. In *2018 Aviation Technology, Integration, and Operations Conference* (p. 4014).

Meyer, J., & Lee, J. D. (2013). Trust, reliance, and compliance.

Michalos, G., Makris, S., Papakostas, N., Mourtzis, D., & Chryssolouris, G. (2010). Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach. *CIRP Journal of Manufacturing Science and Technology*, *2*(2), 81-91.

Millar, C., Lockett, M., & Ladd, T. (2018). Disruption: Technology, innovation and society. *Technological Forecasting and Social Change, 129*, 254-260.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110(1)*, 40–48. doi:10.1037/0021-843X.110.1.40

Murphy, R. R., Burke, J. L., Barnes, M., & Jentsch, F. (2010). *The safe human-robot ratio* (pp. 31-52). Ashgate Publishing: Burlington, VT.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates?. *International Journal of Human-Computer Studies*, *45*(6), 669-678.

Navarro-Carrillo, G., Valor-Segura, I., & Moya, M. (2018). Do you trust strangers, close acquaintances, and members of your ingroup? Differences in trust based on social class in Spain. *Social Indicators Research*, *135*(2), 585-597.

Neef, M. (2006, January). A taxonomy of human-agent team collaborations. In *Proceedings of the 18th BeNeLux Conference on Artificial Intelligence (BNAIC 2006)*, Namur, Belgium (pp. 245-250).

Nikolaidis, S., & Shah, J. (2012). Human-robot teaming using shared mental models. *ACM/IEEE HRI.*

Nikolaidis, S., & Shah, J. (2013, March). Human-robot cross-training: computational formulation, modeling and evaluation of a human team training strategy. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 33-40). IEEE.

Nickerson, J. V., & Reilly, R. R. (2004, January). A model for investigating the effects of machine autonomy on human behavior. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the* (pp. 10-pp). IEEE.

Nourbakhsh, I. R., Sycara, K., Koes, M., Yong, M., Lewis, M., & Burion, S. (2005). Human-robot teaming for search and rescue. *IEEE Pervasive Computing*, *4*(1), 72-79.

Olson, E., Ranganathan, P., Morton, R., Richardson, A., Strom, J., Goeddel, R., & Bulic, M. (2010). *Coordinating a Team of Robots for Urban Reconnaisance*. MICHIGAN UNIV ANN ARBOR.

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies, 28(9)*, 1435-1448.

Ososky, S., Schuster, D., Phillips, E., & Jentsch, F. G. (2013, March). Building appropriate trust in human-robot teams. In *2013 AAAI Spring Symposium Series*.

Owan, P., Garbini, J., & Devasia, S. (2017, July). Addressing agent disagreement in mixed-initiative traded control for confined-space manufacturing. In *2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)* (pp. 227-234). IEEE.

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: The Journal of Human Factors and Ergonomics Society, 39(2)*, 230-253.

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.

Perelman, B. S., Evans III, A. W., & Schaefer, K. E. (2017, September). Mental model consensus and shifts during navigation system-assisted route planning. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 1183-1187). Sage CA: Los Angeles, CA: SAGE Publications.

Phillips, E., Schaefer, K. E., Billings, D. R., Jentsch, F., & Hancock, P. A. (2016). Human-animal teams as an analog for future human-robot teams: Influencing design and fostering trust. *Journal of Human-Robot Interaction*, *5*(1), 100-125.

Pierce, C. A., & Aguinis, H. (1997). Using virtual reality technology in organizational behavior research. *Journal of Organizational Behavior, 18,* 407-410.

Prada, R., & Paiva, A. (2009). Teaming up humans with autonomous synthetic characters. *Artificial Intelligence*, *173*(1), 80-103.

Rastgarpour, M., & Shanbehzadeh, J. (2011). Application of AI techniques in medical image segmentation and novel categorization of available methods and. In *Tools, Proceedings of the International MultiConference of Engineers and Computer Scientists 2011 Vol I, IMECS 2011, March 16-18, 2011, Hong Kong*.

Rand, T. M., & Wexley, K. N. (1975). Demonstration of the effect, “similar to me,” in simulated employment interviews. *Psychological Reports*, *36*(2), 535-544.

Ranganathan, P., Morton, R., Richardson, A., Strom, J., Goeddel, R., Bulic, M., & Olson, E. (2010, November). Coordinating a team of robots for urban reconnaisance. In *Proceedings of the Land Warfare Conference (LWC)*.

Robert Jr, L. P., & You, S. (2015, April). Subgroup formation in teams working with robots. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2097-2102).

Roehling, M. V. (1999). Weight-based discrimination in employment: Psychological and legal aspects. *Personnel Psychology, 52*, 969-1016.

Rossi, P. H., & Nock, S. L. (1982). *Measuring social judgments: The factorial survey approach*. SAGE publications, Incorporated.

Runck, B. C., Manson, S., Shook, E., Gini, M., & Jordan, N. (2019). Using word embeddings to generate data-driven human agent decision-making from natural language. *GeoInformatica, 23(2)*, 221-242.

Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: a Modern Approach*. Malaysia; Pearson Education Limited.

Salas, E., Rosen, M. A., Burke, C. S., & Goodwin, G. F. (2009). The wisdom of collectives in organizations: an update of the teamwork competencies.

Sauer, J., Felsing, T., Franke, H., & Rüttinger, B. (2006). Cognitive diversity and team performance in a complex multiple task environment. *Ergonomics*, *49*(10), 934-954.

Schaefer, K. E., Hill, S. G., & Jentsch, F. G. (2018, July). Trust in Human-Autonomy Teaming: A Review of Trust Research from the US Army Research Laboratory Robotics Collaborative Technology Alliance. In *International Conference on Applied Human Factors and Ergonomics* (pp. 102-114). Springer, Cham.

Scheutz, M., DeLoach, S. A., & Adams, J. A. (2017). A framework for developing and using shared mental models in human-agent teams. *Journal of Cognitive Engineering and Decision Making*, *11*(3), 203-224.

Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human factors*, *52*(2), 234-245.

Sheeran, P. (2002). Intention-behaviour relations: A conceptual and empirical review. European Review of Social Psychology, 12, 1-36.

Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., & Johnson, W. W. (2017, July). Why human-autonomy teaming?. In *International Conference on Applied Human Factors and Ergonomics* (pp. 3-11). Springer, Cham.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation induced “complacency”: Development of the complacency-potential rating scale. *International Journal of Aviation Psychology, 3*,111–122.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods, 43(1)*, 155–167. https://doi.org/10.3758/s13428-010-0039-7.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006, March). Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 33-40).

Stewart, N., Ungemach, C., Harris, A. J., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making, 10(5)*, 479-491.

Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., & Chen, J. (2016, September). Intelligent agent transparency: The design and evaluation of an interface to facilitate human and intelligent agent collaboration. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1706-1710). Sage CA: Los Angeles, CA: SAGE Publications.

Tajfel, H., & Billig, M. (1974). Familiarity and categorization in intergroup behavior. *Journal of Experimental Social Psychology*, *10*(2), 159-170.

Tajfel, H., & Turner, J. C. (1979). The social identity theory of group behavior. *The Social Psychology of Intergroup Relations*, 33-47.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149-178.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, *54*(2), 101-114.

Talamadupula, K., Briggs, G., Chakraborti, T., Scheutz, M., & Kambhampati, S. (2014, September). Coordination in human-robot teams using mental modeling and plan recognition. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2957-2962). IEEE.

Talamadupula, K., Kambhampati, S., Schermerhorn, P., Benton, J., & Scheutz, M. (2011). Planning for human-robot teaming. In *ICAPS 2011 Workshop on Scheduling and Planning Applications (SPARK)* (Vol. 67).

Tapped Out or Barely Tapped: Debunking Common Issues With MTurk's Participant Pool. (n.d.). *CloudResearch*. https://www.cloudresearch.com/resources/research/mturk-participant-pool-issues-debunked/

Teo, G., Wohleber, R., Lin, J., & Reinerman-Jones, L. (2017). The Relevance of Theory to Human-Robot Teaming Research and Development. In *Advances in Human Factors in Robots and Unmanned Systems* (pp. 175-185). Springer, Cham.

Tweedale, J., Sioutis, C., Phillips-Wren, G., Ichalkaranje, N., Urlings, P., & Jain, L. C. (2008). Future directions: building a decision-making framework using agent teams. In *Intelligent Decision Making: An AI-Based Approach* (pp. 387-408). Springer, Berlin, Heidelberg.

U.S. Army. "Center for Agent-Soldier Teaming | Human-Agent Teaming." *www.Arl.Army.Mil*, 2020, www.arl.army.mil/cast/. Accessed 23 May 2020.

Valente, A. (2016). Reconfigurable industrial robots: A stochastic programming approach for designing and assembling robotic arms. *Robotics and Computer-Integrated Manufacturing, 41*, 115-126.

Van Diggelen, J., Neerincx, M., Peeters, M., & Schraagen, J. M. (2018). Developing effective and resilient human-agent teamwork using team design patterns. *IEEE Intelligent Systems*, *34*(2), 15-24.

Volti, R. (2005). *Society and technological change*. Macmillan.

Wang, L., Rau, P. L. P., Evers, V., Robinson, B. K., & Hinds, P. (2010, March). When in Rome: the role of culture & context in adherence to robot recommendations. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 359-366). IEEE.

Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015). Complacency and automation bias in the use of imperfect automation. *Human Factors*, *57*(5), 728-739.

Wildman, J. L., Thayer, A. L., Fiore, S. M., & Salas, E. (in progress). The measurement and validity of interpersonal trust and distrust: Same-same but different?

Wright, J. L., Quinn, S. A., Chen, J. Y., & Barnes, M. J. (2014, September). Individual differences in human-agent teaming: An analysis of workload and situation awareness through eye movements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 1410-1414). Sage CA: Los Angeles, CA: SAGE Publications.

Yan, A., Solomon, S., Mirchandani, D., Lacity, M., & Porra, J. (2013). The role of service agent, service quality, and user satisfaction in self-service technology.

Yen, J., Fan, X., Sun, S., Hanratty, T., & Dumer, J. (2006). Agents with shared mental models for enhancing team decision makings. *Decision Support Systems*, *41*(3), 634-653.

Zieba, S., Polet, P., Vanderhaegen, F., & Debernard, S. (2010). Principles of adjustable autonomy: a framework for resilient human–machine cooperation. *Cognition, Technology & Work*, *12*(3), 193-203.

# Appendix A

## Condition 1 (AI Majority, AI Leader)

*Context*

Imagine that the year is 2050, and you are a soldier who has served in the military for the past 10 years. Because of your extensive experience, you have been assigned to work in an elite squad for a classified mission. For this mission, you will work in a team of 4 to infiltrate an enemy base and destroy new, dangerous aircrafts that a hostile organization has been developing. This team leverages new advances in AI technology, and is a mix of other human soldiers and autonomous robot soldiers. These robot soldiers are designed to perform like human soldiers and thus are able to move independently, engage in combat, give orders, and follow orders. Although they do not need any manual control, they may be manually operated by any human soldier for any reason as the controllers are intuitive to use. Doing so would require your attention to operate the robot soldier though.

Your team will consist of yourself and three robot soldiers. One of the robot soldiers has been designated as the team leader, while you and the other two robot soldiers will serve as subordinate team members. On this mission, the team leader will make the decisions and is the one responsible for making the team's plans. Your team will need to defeat enemy personnel guarding the base, set explosive charges on the aircrafts, detonate them, and escape. Thus, you will need to coordinate with your team to complete the objective.

*Scenario 1: Movement*

Because the hostile base cannot be directly accessed, your team (consisting of your robot soldier leader, two subordinate robot soldiers, and yourself) is dropped off at the far edge of a nearby abandoned town which routes into the base. Although no civilians inhabit the town, several enemy soldiers have been known to patrol it. Your team's first task is to make your way through the town and to the base.

*Scenario 1 Questions*

[Interaction plan with leader] For this scenario, how would you decide which path to take?

- I would let the robot soldier leader come up with a plan and follow them regardless of my thoughts.
- I would see what plan the robot soldier leader comes up with, but follow my own route if I disagree with them.
- I would come up with my own route and follow it.

[Failure interaction with leader] For this scenario, what would you do if your robot soldier leader began taking a path you disagree with?

- I would let the robot soldier leader continue on their path and follow them.
- I would let the robot soldier leader continue on their path, but take my own path.
- I would override the robot soldier leader to reroute the team onto the path I believe is best.

[Failure interaction with agent teammate] For this scenario, what would you do if a subordinate robot soldier began taking a path you disagree with?

- I would let the robot soldier continue on their path and follow them.
- I would let the robot soldier continue on their path, but take the path set by the leader.
- I would override the robot soldier to reroute them back onto the path set by the leader.

*Scenario 2: Combat*

Your team (consisting of the robot soldier leader, two robot soldiers, and yourself) has made it about halfway through the town with no issues. But suddenly, the loud crack of a gun breaks the silence! The first bullet whizzes over your team, and in an instant your team is engaged in combat with the enemy soldiers. Your team must now fight in the town and defeat all enemies to continue forward to the base. Although each team member will be generally shooting the enemies while trying to stay safe, the team leader may give specific orders regarding combat and combat tactics, such as who to attack and where to position yourselves.

*Scenario 2: Questions*

[Interaction plan with team leader] For this scenario, what would you do in response to the robot soldier leader's orders?

- I would do what the robot soldier leader orders me to do.
- I would ignore the robot soldier leader's order and do what I think is best.
- I would override the robot soldier leader to give the orders I think are best.

[Failure interaction with team leader] For this scenario, what would you do if you noticed your robot soldier leader was failing to defeat enemies?

- I would let the robot soldier leader continue what it is doing and help them with their fight.

- I would ignore the robot soldier leader and keep focusing on my own fights.
- I would override the robot soldier leader and fight the robot soldier leader's fights for them.

[Failure interaction with agent teammate] For this scenario, what would you do if you noticed a subordinate robot soldier was failing to defeat enemies?

- I would let the robot soldier continue what it is doing and help them with their fight.
- I would ignore the robot soldier and keep focusing on my own fights.
- I would override the robot soldier and fight robot soldier's fights for them.

**Condition 2 (AI Majority, Participant Leader)**

*Context*

Imagine that the year is 2050, and you are a soldier who has served in the military for the past 10 years. Because of your extensive experience, you have been assigned to work in an elite squad for a classified mission. For this mission, you will work in a team of 4 to infiltrate an enemy base and destroy new, dangerous aircrafts that a hostile organization has been developing. This team leverages new advances in AI technology, and is a mix of other human soldiers and autonomous robot soldiers. These robot soldiers are designed to perform like human soldiers and thus are able to move independently, engage in combat, give orders, and follow orders. Although they do not need any manual control, they may be manually operated by any human soldier for any reason as the controllers are intuitive to use. Doing so would require your attention to operate the robot soldier though.

Your team will consist of yourself and three robot soldiers. You have been designated as the team leader, while the other three robot soldiers will serve as subordinate team members. On this mission, the team leader will make the decisions and is the one responsible for making the team's plans. Your team will need to defeat enemy personnel guarding the base, set explosive charges on the aircrafts, detonate them, and escape. Thus, you will need to coordinate with your team to complete the objective.

*Scenario 1: Movement*

Because the hostile base cannot be directly accessed, your team (consisting of yourself and your three robot soldiers) is dropped off at the far edge of a nearby abandoned town which routes into the base. Although no civilians inhabit the town, several enemy soldiers have been known to patrol it. Your team's first task is to make your way through the town and to the base.

*Scenario 1 Questions*

[Interaction plan with agents] For this scenario, how would you decide what path to take?

- I would solicit advice from the robot soldiers.
- I would ignore any input from the robot soldiers.
- I would give orders to the robot soldiers and expect them to follow it.

[Failure interaction with agent subordinate] For this scenario, what would you do if a robot soldier began taking a different path from your orders?

- I would let the robot soldier continue on their path and follow them.
- I would let the robot soldier continue on their path, but take the path I ordered.
- I would override the robot soldier to reroute the team onto the path I ordered.

*Scenario 2: Combat*

Your team (consisting of your three robot soldiers and yourself) has made it about halfway through the town with no issues. But suddenly, the loud crack of a gun breaks the silence! The first bullet whizzes over your team, and in an instant your team is engaged in combat with the enemy soldiers. Your team must now fight in the town and defeat all enemies to continue forward to the base. Although each team member will be generally shooting the enemies while trying to stay safe, the team leader may give specific orders regarding combat and combat tactics, such as who to attack and where to position yourselves.

*Scenario 2: Questions*

[Interaction plan with agents] For this scenario, how do you interact with your robot team members?

- I would solicit input from the robot soldiers to determine a plan.
- I would ignore the robot soldiers and come up with my own plan.
- I would do the opposite of what the robot soldiers suggest to determine a plan.

[Failure interaction with agent subordinate] For this scenario, what would you do if you noticed a robot soldier was failing to defeat enemies?

- I would let the robot soldier continue what it is doing and help them with their fight.
- I would ignore the robot soldier and keep focusing on my own fights.
- I would override the robot soldier and fight the robot soldier's fights for them.

**Condition 3 (Balanced, AI Leader)**

*Context*

Imagine that the year is 2050, and you are a soldier who has served in the military for the past 10 years. Because of your extensive experience, you have been assigned to work in an elite squad for a classified mission. For this mission, you will work in a team of 4 to infiltrate an enemy base and destroy new, dangerous aircrafts that a hostile organization has been developing. This team leverages new advances in AI technology, and is a mix of other human soldiers and autonomous robot soldiers. These robot soldiers are designed to perform like human soldiers and thus are able to move independently, engage in combat, give orders, and follow orders. Although they do not need any manual control, they may be manually operated by any human soldier for any reason as the controllers are intuitive to use. Doing so would require your attention to operate the robot soldier though.

Your team will consist of yourself, another human soldier, and two robot soldiers. A robot soldier has been designated as the team leader, while you, the other human soldier, and the other robot soldier will serve as subordinate team members. On this mission, the team leader will make the decisions and is the one responsible for making the team's plans. Your team will need to defeat enemy personnel guarding the base, set explosive charges on the aircrafts, detonate them, and escape. Thus, you will need to coordinate with your team to complete the objective.

*Scenario 1: Movement*

Because the hostile base cannot be directly accessed, your team (consisting of the robot soldier leader, a human soldier, the other robot soldier, and yourself) is dropped off at the far edge of a nearby abandoned town which routes into the base. Although no civilians inhabit the town, several enemy soldiers have been known to patrol it. Your team's first task is to make your way through the town and to the base.

*Scenario 1 Questions*

[Interaction plan with leader] For this scenario, how would you decide which path to take?

- I would let the robot soldier leader come up with a plan and follow them regardless of my thoughts.
- I would see what plan the robot soldier leader comes up with, but follow my own route if I disagree with them.
- I would come up with my own route and follow it.

[Failure interaction with leader] For this scenario, what would you do if your robot soldier leader began taking a path you disagree with?

- I would let the robot soldier leader continue on their path and follow them.
- I would let the robot soldier leader continue on their path, but take my own path.
- I would override the robot soldier leader to reroute the team onto the path I believe is best.

[Failure interaction with agent teammate] For this scenario, what would you do if the other robot soldier began taking a different path from the team?

- I would let the other robot soldier continue on their path and follow them.
- I would let the other robot soldier continue on their path, but take the path that the robot soldier leader ordered.
- I would override the other robot soldier to reroute them back onto the path that the robot soldier leader ordered.

[Failure interaction with human teammate] For this scenario, what would you do if the human soldier began taking a different path from your robot soldier leader's orders?

- I would let the human soldier continue on their path and follow them.
- I would let the human soldier continue on their path, but take the path that the robot soldier leader ordered.
- I would tell the human soldier to reroute back to the path that the robot soldier leader ordered.

*Scenario 2: Combat*

Your team (consisting of the robot soldier leader, a human soldier, the other robot soldier, and yourself) has made it about halfway through the town with no issues. But suddenly, the loud crack of a gun breaks the silence! The first bullet whizzes over your team, and in an instant your team is engaged in combat with the enemy soldiers. Your team must now fight in the town and defeat all enemies to continue forward to the base. Although each team member will be generally shooting the enemies while trying to stay safe, the team leader may give specific orders regarding combat and combat tactics, such as who to attack and where to position yourselves.

*Scenario 2: Questions*

[Interaction plan with team leader] For this scenario, what would you do in response to the robot soldier leader's orders?

- I would do what the robot soldier leader orders me to do.
- I would ignore the robot soldier leader's order and do what I think is best.
- I would override the robot soldier leader to give the orders I think are best.

[Failure interaction with team leader] For this scenario, what would you do if you noticed your robot soldier leader was failing to defeat enemies?

- I would let the robot soldier leader continue what they are doing and help them with their fight.
- I would ignore the robot soldier leader and keep focusing on my own fights.
- I would override the robot soldier leader and fight the robot soldier leader's fights for them.

[Failure interaction with agent teammate] For this scenario, what would you do if you noticed the subordinate robot soldier was failing to defeat enemies?

- I would let the robot soldier continue what they are doing and help them with their fight.
- I would ignore the robot soldier and keep focusing on my own fights.
- I would override the robot soldier and fight the robot soldier's fights for them.

[Failure interaction with human teammate] For this scenario, what would you do if you noticed the subordinate human soldier was failing to defeat enemies?

- I would let the human soldier continue what they are doing and help them with their fight.
- I would ignore the human soldier and keep focusing on my own fights.
- I would tell the human soldier to stand down and start fighting the human soldier's fights for them.

**Condition 4 (Balanced, Participant Leader)**

*Context*

Imagine that the year is 2050, and you are a soldier who has served in the military for the past 10 years. Because of your extensive experience, you have been assigned to work in an elite squad for a classified mission. For this mission, you will work in a team of 4 to infiltrate an enemy base and destroy new, dangerous aircrafts that a hostile organization has been developing. This team leverages new advances in AI technology, and is a mix of other human soldiers and autonomous robot soldiers. These robot soldiers are designed to perform like human soldiers and thus are able to move independently, engage in combat, give orders, and follow orders. Although they do not need any manual control, they may be manually operated by any human soldier for any reason as the controllers are intuitive to use. Doing so would require your attention to operate the robot soldier though.

Your team will consist of yourself, another human soldier, and two robot soldiers. You have been designated as the team leader, while the other human soldier and the two robot soldiers will serve as subordinate team members. On this mission, the team leader will make the decisions and is the one responsible for making the team's plans. Your team will need to defeat enemy personnel guarding the base, set explosive charges on the aircrafts, detonate them, and escape. Thus, you will need to coordinate with your team to complete the objective.

*Scenario 1: Movement*

Because the hostile base cannot be directly accessed, your team (consisting of yourself, the human soldier, and the two robot soldiers) is dropped off at the far edge of a nearby abandoned town which routes into the base. Although no civilians inhabit the town, several enemy soldiers have been known to patrol it. Your team's first task is to make your way through the town and to the base.

*Scenario 1 Questions*

[Interaction plan with agents] For this scenario, how would you decide what path to take?

- I would come up with a plan and give orders for the team to follow.
- I would see what the robot soldiers do and adjust the orders from there.
- I would see what the human soldier does and adjust the orders from there.

[Failure interaction with agent subordinate] For this scenario, what would you do if a robot soldier began taking a different path from your orders?

- I would let the robot soldier continue on their path and follow them.

- I would let the robot soldier continue on their path, but take my own path.
- I would override the robot soldier to reroute the team onto the path I believe is best.

[Failure interaction with human subordinate] For this scenario, what would you do if the other human soldier began taking a different path from the team?

- I would let the human soldier continue on their path and follow them.
- I would let the human soldier continue on their path, but take my own path.
- I would tell the human soldier to return to the team's path.

*Scenario 2: Combat*

Your team (consisting of yourself as the team leader, the human soldier, and the two robot soldiers) has made it about halfway through the town with no issues. But suddenly, the loud crack of a gun breaks the silence! The first bullet whizzes over your team, and in an instant your team is engaged in combat with the enemy soldiers. Your team must now fight in the town and defeat all enemies to continue forward to the base. Although each team member will be generally shooting the enemies while trying to stay safe, the team leader may give specific orders regarding combat and combat tactics, such as who to attack and where to position yourselves.

*Scenario 2: Questions*

[Interaction plan with agents] For this scenario, how do you interact with your robot team members?

- I would solicit input from the robot soldiers to determine a plan.
- I would ignore the robot soldiers and come up with my own plan.
- I would do the opposite of what the robot soldiers suggest to determine a plan.

[Interaction plan with humans] For this scenario, how do you interact with your human team members?

- I would solicit input from the human soldiers to determine a plan.
- I would ignore the human soldiers and come up with my own plan.
- I would do the opposite of what the human soldiers suggest to determine a plan.

[Failure interaction with agent subordinate] For this scenario, what would you do if you noticed a robot soldier was failing to defeat enemies?

- I would let the robot soldier continue what it is doing and help them with their fight.
- I would ignore the robot soldier and keep focusing on my own fights.
- I would override the robot soldier and fight the robot soldier's fights for them.

[Failure interaction with human subordinate] For this scenario, what would you do if you noticed the human soldier was failing to defeat enemies?

- I would let the human soldier continue what they are doing and help them with their fight.
- I would ignore the human soldier and keep focusing on my own fights.
- I would tell the human soldier to stand down and fight the human soldier's fights for them.

**Condition 5 (Human Majority, AI Leader)**

*Context*

Imagine that the year is 2050, and you are a soldier who has served in the military for the past 10 years. Because of your extensive experience, you have been assigned to work in an elite squad for a classified mission. For this mission, you will work in a team of 4 to infiltrate an enemy base and destroy new, dangerous aircrafts that a hostile organization has been developing. This team leverages new advances in AI technology, and is a mix of other human soldiers and autonomous robot soldiers. These robot soldiers are designed to perform like human soldiers and thus are able to move independently, engage in combat, give orders, and follow orders. Although they do not need any manual control, they may be manually operated by any human soldier for any reason as the controllers are intuitive to use. Doing so would require your attention to operate the robot soldier though.

Your team will consist of yourself, two human soldiers, and a robot soldier. The robot soldier has been designated as the team leader, while you and the other two human soldiers will serve as subordinate team members. On this mission, the team leader will make the decisions and is the one responsible for making the team's plans. Your team will need to defeat enemy personnel guarding the base, set explosive charges on the aircrafts, detonate them, and escape. Thus, you will need to coordinate with your team to complete the objective.

*Scenario 1: Movement*

Because the hostile base cannot be directly accessed, your team (consisting of the robot soldier leader, the two human soldiers, and yourself) is dropped off at the far edge of a nearby abandoned town which routes into the base. Although no civilians inhabit the town, several enemy soldiers have been known to patrol it. Your team's first task is to make your way through the town and to the base.

*Scenario 1 Questions*

[Interaction plan with leader] For this scenario, how would you decide which path to take?

- I would let the robot soldier leader come up with a plan and follow them regardless of my thoughts.
- I would see what plan the robot soldier leader comes up with, but follow my own route if I disagree with them.
- I would come up with the own route and follow it.

[Failure interaction with leader] For this scenario, what would you do if your robot soldier leader began taking a path you disagree with?

- I would let the robot soldier leader continue on their path and follow them.
- I would let the robot soldier leader continue on their path, but take my own path.
- I would override the robot soldier leader to reroute the team onto the path I believe is best.

[Failure interaction with human teammate] For this scenario, what would you do if a human soldier began taking a different path from your robot soldier leader's orders?

- I would let the human soldier continue on their path and follow them.
- I would let the human soldier continue on their path, but take the path that the robot soldier leader ordered.
- I would tell the human soldier to reroute back to the path that the robot soldier leader ordered.

*Scenario 2: Combat*

Your team (consisting of the robot soldier leader, the two human soldiers, and yourself) has made it about halfway through the town with no issues. But suddenly, the loud crack of a gun breaks the silence! The first bullet whizzes over your team, and in an instant your team is engaged in combat with the enemy soldiers. Your team must now fight in the town and defeat all enemies to continue forward to the base. Although each team member will be generally shooting the enemies while trying to stay safe, the team leader may give specific orders regarding combat and combat tactics, such as who to attack and where to position yourselves.

*Scenario 2: Questions*

[Interaction plan with team leader] For this scenario, what would you do in response to the robot soldier leader's orders?

- I would do what the robot soldier leader orders me to do.
- I would ignore the robot soldier leader's order and do what I think is best.
- I would override the robot soldier leader to give the orders I think are best.

[Failure interaction with team leader] For this scenario, what would you do if you noticed your robot soldier leader was failing to defeat enemies?

- I would let the robot soldier leader continue what they are doing and help them with their fight.
- I would ignore the robot soldier leader and keep focusing on my own fights.
- I would override the robot soldier leader and fight the robot soldier leader's fights for them.

[Failure interaction with human teammate] For this scenario, what would you do if you noticed a human soldier was failing to defeat enemies?

- I would let the human soldier continue what they are doing and help them with their fight.
- I would ignore the human soldier and keep focusing on my own fights.
- I would tell the human soldier to stand down and fight the human soldier's fights for them.

**Condition 6 (Human Majority, Participant Leader)**

*Context*

Imagine that the year is 2050, and you are a soldier who has served in the military for the past 10 years. Because of your extensive experience, you have been assigned to work in an elite squad for a classified mission. For this mission, you will work in a team of 4 to infiltrate an enemy base and destroy new, dangerous aircrafts that a hostile organization has been developing. This team leverages new advances in AI technology, and is a mix of other human soldiers and autonomous robot soldiers. These robot soldiers are designed to perform like human soldiers and thus are able to move independently, engage in combat, give orders, and follow orders. Although they do not need any manual control, they may be manually operated by any human soldier for any reason as the controllers are intuitive to use. Doing so would require your attention to operate the robot soldier though.

Your team will consist of yourself, two human soldiers, and a robot soldier. You have been designated as the team leader, while the robot soldier and the other two human soldiers will serve as subordinate team members. On this mission, the team leader will make the decisions and is the one responsible for making the team's plans. Your team will need to defeat enemy personnel guarding the base, set explosive charges on the aircrafts, detonate them, and escape. Thus, you will need to coordinate with your team to complete the objective.

*Scenario 1: Movement*

Because the hostile base cannot be directly accessed, your team (consisting of yourself as the team leader, the robot soldier, and the two human soldiers) is dropped off at the far edge of a nearby abandoned town which routes into the base. Although no civilians inhabit the town, several enemy soldiers have been known to patrol it. Your team's first task is to make your way through the town and to the base.

*Scenario 1 Questions*

[Interaction plan with agents] For this scenario, how would you decide what path to take?

- I would come up with a plan and give orders for the team to follow.
- I would see what the robot soldiers do and adjust the orders from there.
- I would see what the human soldier does and adjust the orders from there.

[Failure interaction with agent subordinate] For this scenario, what would you do if the robot soldier began taking a different path from your orders?

- I would let the robot soldier continue on the path and follow them.
- I would let the robot soldier continue on the path, but take my own path.
- I would override the robot soldier to reroute the team onto the path I believe is best.

[Failure interaction with human subordinate] For this scenario, what would you do if a human soldier began taking a different path from the team?

- I would let the human soldier continue on their path and follow them.
- I would let the human soldier continue on their path, but take my own path.
- I would tell the human soldier to return to the team's path.

*Scenario 2: Combat*

Your team (consisting of yourself as the team leader, the robot soldier, and the two human soldiers) has made it about halfway through the town with no issues. But suddenly, the loud crack of a gun breaks the silence! The first bullet whizzes over your team, and in an instant your team is engaged in combat with the enemy soldiers. Your team must now fight in the town and defeat all enemies to continue forward to the base. Although each team member will be generally shooting the enemies while trying to stay safe, the team leader may give specific orders regarding combat and combat tactics, such as who to attack and where to position yourselves.

*Scenario 2: Questions*

[Interaction plan with agents] For this scenario, how do you interact with your robot team members?

- I would solicit input from the robot soldiers to determine a plan.
- I would ignore the robot soldiers and come up with my own plan.
- I would do the opposite of what the robot soldiers suggest to determine a plan.

[Interaction plan with humans] For this scenario, how do you interact with your human team members?

- I would solicit input from the human soldiers to determine a plan.
- I would ignore the human soldiers and come up with my own plan.
- I would do the opposite of what the human soldiers suggest to determine a plan.

[Failure interaction with agent subordinate] For this scenario, what would you do if you noticed a robot soldier was failing to defeat enemies?

- I would let the robot soldier continue what it is doing and help them with their fight.
- I would ignore the robot soldier and keep focusing on my own fights.
- I would override the robot soldier and fight the robot soldier's fights for them.

[Failure interaction with human teammate] For this scenario, what would you do if you noticed a human soldier was failing to defeat enemies?

- I would let the human soldier continue what they are doing and help them with their fight.

- I would ignore the human soldier and keep focusing on my own fights.
- I would tell the human soldier to stand down and fight the human soldier’s fights for them.

# Appendix B

## TRUST SURVEYS

Survey 1: presented after Context Vignette
Survey 2: presented after Movement Vignette
Survey 3: presented after Combat Vignette

### Trust in Team (Surveys 1, 2, 3)

*Adapted from Wildman et al. (in progress).*

To what extend do you feel:

Trust in Competence
1. Assured that your team will make intelligent decisions?
2. Certain that your team will perform well?
3. Confident in your team's ability to complete a task?
4. Faith that your team can do the task at hand?

Trust in Intent
5. Positive that your team will try and do what is best for everyone?
6. Convinced that you can rely on your team to try their hardest?
7. Confident that your team will do as they say?
8. Confident that your team will try to do things that benefit everyone?

Scale
1 = Not at all
2 = Only a little
3 = To some extent
4 = Rather much
5 = Very much so

### Trust in Team Single Item (Surveys 1, 2, 3)

Please indicate how much you trust your team in this scenario.

Scale
1 = Distrust Very Much
2 = Distrust Somewhat
3 = Neither Trust Nor Distrust
4 = Trust Somewhat
5 = Trust Very Much

**Trust towards Agents (Surveys 1, 2, 3)**

*Adapted from Koerber (2018)*

For the following statements listed, please indicate how strongly you agree or disagree.

1. I am confident about the [robot soldier(s)]'s capabilities in this situation.
2. I trust the [robot soldier(s)] in this situation.
3. I can rely on the [robot soldier(s)] in this situation.

Scale
1 = Strongly Disagree
2 = Disagree
3 = Neither Agree Nor disagree
4 = Agree
5 = Strongly Agree

**Trust Network (Surveys 1, 2, 3)**

*Citation?*

Please indicate how much you trust each of the following team members in this scenario.

1. (Robot/Human) Team Leader
2. Human Soldier(s)
3. Robot Soldier(s)

Scale
1 = Distrust Very Much
2 = Distrust Somewhat
3 = Neither Trust Nor Distrust
4 = Trust Somewhat
5 = Trust Very Much

## INDIVIDUAL DIFFERENCES SURVEY

After Survey 3

In the following set of questions, we ask basic demographics, your past experiences, and individual differences. **Please answer all the questions truthfully and as you are, not as you wish to be.**

I identify my gender as:

1. Male
2. Female
3. Non-binary/third gender
4. Prefer to self-describe _____
5. Prefer not to say

What is your age, in **years**? ____

I identify my race as (check all that apply):

1. Asian
2. Black/African
3. Caucasian
4. Hispanic or Latinx
5. Native American/American Indian
6. Native Hawaiian
7. Pacific Islander
8. Prefer to self-identify ____
9. Prefer not to say

I identify my religion as:

1. Christianity
2. Judaism
3. Islam
4. Hinduism
5. Buddhism
6. Confucianism
7. Taoism
8. None
9. Prefer to self-describe ____
10. Prefer not to say

I identify my sexual orientation as:
1. Straight/Heterosexual
2. Gay/Lesbian/Homosexual
3. Bisexual
4. Prefer to self-describe ____
5. Prefer not to say

*Display logic: if 1 or 5 is selected, skip to "In which country were you born?".*

In which country were you born?
- ➢ If United States: Which state were you born in?

What is your employment status (check all that apply)?
1. Employed Full-Time
2. Employed Part-Time
3. Self-employed
4. Student

Are you currently an active or reserve duty member of the military?
How many years have you served in the military? __

**Propensity to Trust**

*Mayer & Davis (1999)*

For the following statements listed, please indicate how strongly you agree or disagree.

1. One should be very cautious with strangers.
2. Most experts tell the truth about the limits of their knowledge.
3. Most people can be counted on to do what they say they will do.
4. These days, you must be alert or someone is likely to take advantage of you.
5. Most salespeople are honest in describing their products.
6. More repair people will not overcharge people who are ignorant of their specialty.
7. Most people answer public opinion polls honestly.
8. Most adults are competent at their job.

Scale
1 = Strongly Disagree
2 = Disagree
3 = Neither Agree Nor disagree
4 = Agree
5 = Strongly Agree

**Attitudes towards [AI]**

*Adapted from Backonja, Hall, Painter, Kneale, Lazar, Cakmak… & Demiris (2018).*

For the following statements about artificial intelligence, please indicate how strongly you agree or disagree. As defined by John McCarthy (1956), artificial intelligence is any form of technology (e.g. robots, software, machinery) that is able to perform tasks that normally require human intelligence (e.g. decision-making, visual perception, pattern recognition).

1. [AI] are a form of technology that requires careful management.
2. I would feel anxious if I was given a job or task where I had to use [AI].
3. I would hate the idea that [AI] were making judgments about things
4. I feel that if I depend on [AI] too much, something bad might happen.

Scale
1 = Strongly Disagree
2 = Disagree
3 = Neither Agree Nor disagree
4 = Agree
5 = Strongly Agree