AFRL-AFOSR-VA-TR-2023-0220

Evaluating factors that affect trust calibration: the influence of trust strategy and risk

Shaw, Tyler
GEORGE MASON UNIVERSITY
4400 UNIVERSITY DR
FAIRFAX, VA, 22030
USA

12/01/2022
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**	
20221201	Final	**START DATE** 20151201	**END DATE** 20190531

4. TITLE AND SUBTITLE
Evaluating factors that affect trust calibration: the influence of trust strategy and risk

5a. CONTRACT NUMBER	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
	FA9550-16-1-0023	61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Tyler Shaw

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
GEORGE MASON UNIVERSITY
4400 UNIVERSITY DR
FAIRFAX, VA 22030
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	AFRL/AFOSR RTA2	AFRL-AFOSR-VA-TR-2023-0220

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Recent years have seen a shift away from “automated” systems that support human performance towards “autonomous” systems that are essentially self-governing. Because of the collaborative and often interdependent nature of human-machine performance, issues surrounding human-machine trust have become more important than ever. This effort was designed to explore issues relating to trust calibration that influence the way in which operators interact with systems. This report summarizes the major research activities, study results, and research accomplishments associated with the grant entitled “Evaluating factors that affect trust calibration: the influence of trust strategy and risk.” This is also the final report. My research team and I have coordinated several different research thrusts on trust calibration, situation-specific trust, and human-machine teaming. From the research, we have found that 1) we can alter the way in which trust is allocated by employing mitigation techniques that prevent over-trusting, 2) by using dynamic methods of changing risk, in a laboratory setting, you get differential effects of trust, and 3) improving the social relationship between humans and autonomy can lead to superior human-machine performance outcomes. Each of these three research thrusts, and the major accomplishments of the grant, is discussed in detail.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		
U	U	U	UU	19

19a. NAME OF RESPONSIBLE PERSON	**19b. PHONE NUMBER** *(Include area code)*
RICHARD RIECKEN	696-9736

Evaluating factors that affect trust calibration: the influence of trust strategy and risk

PI: Tyler Shaw, tshaw4@gmu.edu, 703-993-5187

Reporting Period: 12/01/2015- 05/31/2019

Recent years have seen a shift away from "automated" systems that support human performance towards "autonomous" systems that are essentially self-governing. Because of the collaborative and often interdependent nature of human-machine performance, issues surrounding human-machine trust have become more important than ever. This effort was designed to explore issues relating to trust calibration that influence the way in which operators interact with systems. This report summarizes the major research activities, study results, and research accomplishments associated with the grant entitled "Evaluating factors that affect trust calibration: the influence of trust strategy and risk." This is also the final report of the project. My research team and I have coordinated several different research thrusts on trust calibration, situation-specific trust, and human-machine teaming. From the research, we have found that 1) we can alter the way in which trust is allocated by employing mitigation techniques that prevent over-trusting, 2) by using dynamic methods of changing risk, in a laboratory setting, you get differential effects of trust, and 3) improving the social relationship between humans and autonomy can lead to superior human-machine performance outcomes. Each of these three research thrusts, and the major accomplishments of the grant, is discussed in detail.

Table of Contents

List of Figures

List of Tables

Research and Educational Activities

This section outlines the research and educational activities that were facilitated by this grant on calibrating trust in unmanned aerial vehicle (UAV) simulations.

1. Introduction

While advances in technology in the middle of the twentieth century led to the implementation of automation into human-machine interaction, a similar trend is occurring today with the explosion of "autonomous" systems. While automation is designed to support decision making and offload tasks that were originally intended to be performed by the operator, autonomy is a set of capabilities that are "self-governing", within programmed constraints (Defense Science Board, 2012). It has long been known that there are unintended consequences in dealing with automation (see Parasuraman & Riley, 1997 for a review), but one factor that is gaining much more traction recently is the notion of operator trust.

It has been assumed for some time that trust is a construct worthy of empirical study in HF (see Parasuraman & Riley, 1997 for an early review), but it should be noted that not all researchers agree (e.g. Dekker & Woods, 2002). While trust is largely a social psychological concept, it has been used in human factors to partly characterize human-machine partnerships. Formally defined, trust is an attitude that an agent or system will help achieve an operator's goals in an uncertain or vulnerable situation. (Lee & See, 2004).Whether human-human trust is qualitatively similar or different from human-machine trust is an issue that is still being explored (e.g. Madhavan, 2007), but the empirical work to date has shown that trust is important because it dictates a particular strategy for interacting with automation and autonomous systems. In that respect, trust operates as an intervening variable between system capabilities and performance.

Since trust may have an indirect impact on the appropriate use of autonomy, trust is thus operationalized in terms of human performance. It is usually spoken about in terms of *reliance* on an automated or autonomous system (e.g. Lee & Moray, 1994; Parasuraman & Riley, 1997) and that an appropriate, *well-calibrated* level of reliance could contribute to keeping performance at near optimal levels. Very high levels of trust in an autonomous system that is not 100% reliable can result in a strategy that yields complacency (Moray & Inagaki, 2000). Low levels of trust can result in disuse, which means that functions that could very well be carried out successfully by the automated or autonomous agent are being neglected (Parasuraman et al., 1997).

Thus, trust calibration, that is, matching appropriate levels of operator trust with appropriate levels of autonomous system reliability, has become an important area of study in recent years. It is thought that the calibration of trust will result in more appropriate reliance and reduced error. Operators will experience reduced cognitive load based on interfaces that require minimal supervision. This frees up their attention to perform other tasks. With calibrated trust, operators will be quicker to perceive and respond to unexpected events in rapidly changing environments. Then it may be possible to have greater number of autonomous systems effectively managed by a reduced number of operators.

During the past 3 years, we have conducted three major efforts involving:

- Examining how the manner in which trust is allocated can influence the way in which operators interact with systems.
- Exploring an often understudied aspect of trust, operator risk, and it's subsequent influence on trust strategy
- Exploring potential interaction structures for human-machine teaming

In the following part of the report each study will be described in detail.

2. Effort #1: The influence of trust strategy on monitoring behavior of multiple unmanned aerial vehicle

2.1. Motivation

The purpose of the first effort was to examine whether people apply their trust generally or specifically when supervising multiple autonomous agents. More precisely, do operators apply a system-wide trust (SWT) or component specific trust (CST) while monitoring multiple agents within a system. Under SWT, the treatment of each agent is dependent on the performance of other agents within the system; every agent in the system contributes to a general assessment of trust. Conversely, when the agents within a system are treated independently from each other, a CST strategy is employed. A system-wide trust strategy may be problematic if it results in disuse of accurate agents within the system (Keller & Rice, 2010). A recent study has demonstrated that a single unreliable agent can create a "pull-down" effect such that the rate of operator trust in other functionally similar agents is lowered (Keller & Rice, 2010). The pull-down effect from the unreliable aid provided support for the prediction that people tend to apply a system-wide trust strategy.

One important aspect of SWT research is the identification of interventions that can reduce the pull-down effect and support a CST strategy. One factor, performance feedback, has previously been demonstrated to reduce the number of suboptimal automation usage decisions (Beck, Dzindolet, & Pierce, 2007). Following the initial SWT study, further research explored the possibility that mitigation strategies, (i.e. providing knowledge of system accuracy and performance feedback) could support component specific trust (Rice & Geels, 2010). The results indicated that even though overall response accuracy was improved when the aides were known to be perfectly reliable, SWT remained the dominant strategy (Geels-Blair et al, 2013). These studies provided strong support for the prevalence of SWT, however, SWT and the mitigating effect of system performance information had not been studied in realistic settings such as collaboration with autonomous agents in a multi-agent system.

The study described in this report addressed this issue by examining SWT in a supervisory control setting that more closely resembles a real-world multiple unmanned aerial vehicle control scenario. Using a supervisory control UAV simulation in which the reliability of the UAVs were manipulated, it was predicted that a single inaccurate vehicle would reduce trust and reliance for all vehicles equally. In addition, we predicted that mitigation strategies, in the form of vehicle accuracy and performance feedback, would partially support CST as they had in previous studies. This would be observed as a separation of the trust ratings of the accurate

vehicles relative to the inaccurate vehicle. It was predicted that performance information and feedback would uniformly increase trust and reliance when all vehicles were perfectly accurate.

2.2 Experiment

This experiment examined SWT in a supervisory control setting that more closely resembled a real-world multiple unmanned aerial vehicle control scenario. Using a supervisory control UAV simulation in which the reliability of the UAVs were manipulated, we predicted that a single inaccurate vehicle would reduce trust and reliance for all vehicles equally, even though the other vehicles were perfectly reliable. In addition, we predicted that mitigation strategies, in the form of transparency of vehicle accuracy and providing performance feedback, would foster a more appropriate component specific trust strategy. This would be observed as a separation of the trust ratings of the accurate vehicles relative to the inaccurate vehicle. It is also expected that performance information would also uniformly increase trust and reliance when all vehicles were perfectly accurate. We also expect that the same patterning of results will be present in accuracy, verification behaviors, response time, and subjective ratings of perceived reliance.

We collected data on one hundred and sixty one students. This experiment employed a 2 x 2 between-subjects design with vehicle accuracy (Perfect-100%, Imperfect-70%) and performance information (Informed, Uninformed). Vehicle accuracy was a manipulation of one of the four UAVs correct identification rate—participants were assigned to a condition where one of the vehicles was only 70% accurate. The three remaining vehicles (UAVs 2-4) were 100% reliable in both conditions. The Performance Information variable also had two levels. Participants in the informed condition were provided the vehicle accuracy rates prior to beginning the experiment and also received performance feedback after each response. Those in the uninformed condition were not provided vehicle accuracy rates or performance feedback during the experiment. A desktop computer was utilized to run a JAVA based program titled, “Research Environment for Supervisory Control of Heterogeneous Unmanned Vehicles” (RESCHU) that provides a customizable environment to interact with multiple unmanned vehicles. The majority of the RESCHU interface is used for a map display which provides continuously updated locations of vehicles and targets (See Fig. 1). The left hand side of the display provides current vehicle information, imagery, and buttons that allow for interaction with specific vehicles. For this study, RESCHU was configured for participants to supervise the operation of four unmanned aerial vehicles as they flew to unidentified targets. The targets appeared at random and the aircraft flew to targets without input from participants. The program was also configured to employ an “automated target recognition” system that identified the vehicles in photographs as “enemy” or “friendly”. Participants were given the opportunity to view photographs of the target vehicles for up to 2.5 seconds if they decided to do so. Participants also had the option to simply comply with the automation blindly and not view the photographs.

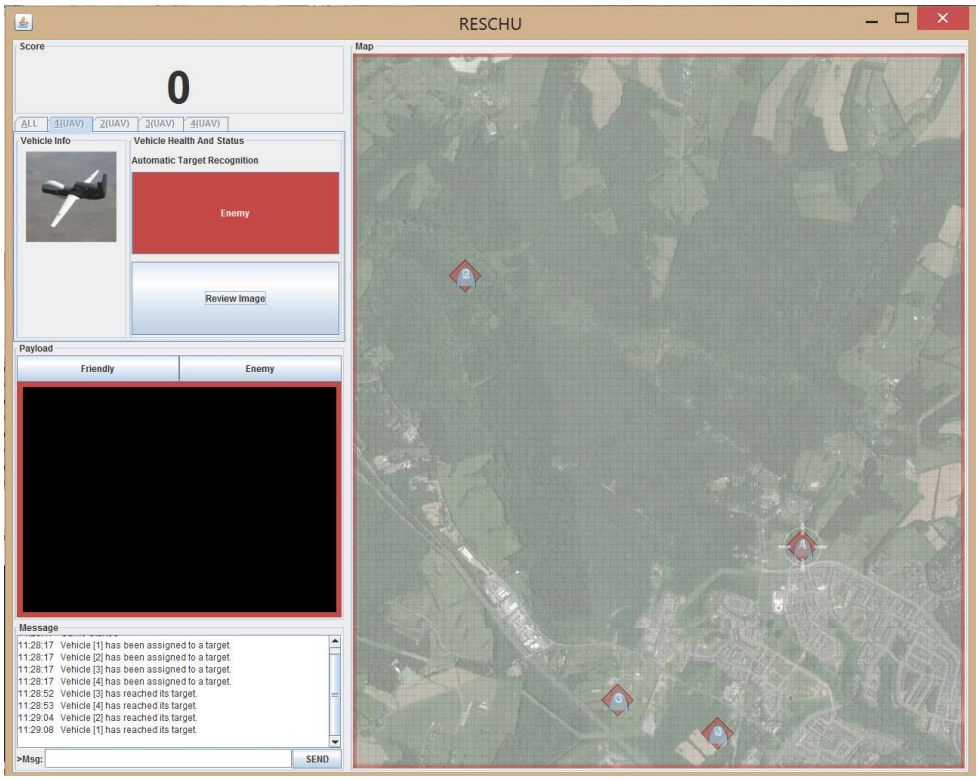

Figure 1 RESCHU interface showing the map display (right) and the automation recommendation and the decision panel (left).

The results of experiment 1 pertaining to performance and trust revealed that that participants were significantly more accurate with the perfectly reliable UAVs (#2-4) than with the unreliable UAV (UAV#1). More interestingly, the performance and trust results revealed that when information was provided, participants were able to better calibrate their performance and trust (see figures 2 and 3, respectively). This result is consistent with evidence that suggests that providing system transparency can facilitate better affective and performance outcomes. It is noteworthy that, as was the case with the feedback manipulation, the information manipulation did not eliminate the SWT effect entirely, especially with regard to performance. While performance in the informed and imperfect condition was slightly higher than performance in the uninformed and imperfect condition (indicating less of a "pull down" effect), a similar patterning of results occurred across those two conditions. Where the information manipulation made the biggest impact is with regard to trust. It was observed that in the conditions in which the UAV was imperfect but information was provided, participants were better able to calibrate their trust to the *actual* reliability of the system (Figure 3). This was observed with subjective trust, and it was not observed in the uninformed and imperfect condition. In the latter condition, subjective trust and verification behavior remained relatively uniform across all of the UAVs. This finding suggests that the information condition did help users adopt more of a component specific trust strategy, though it didn't translate to better performance outcomes.

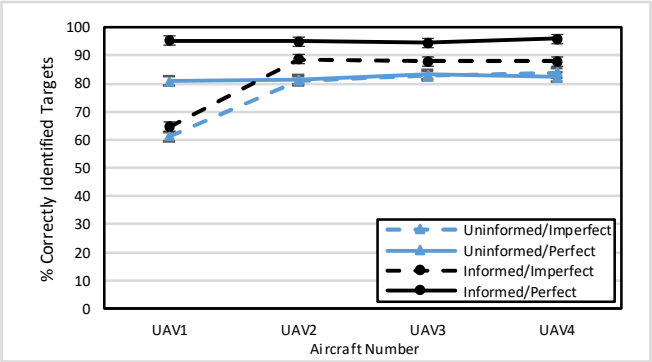

Figure 2. Accuracy Rate with SEM error bars.

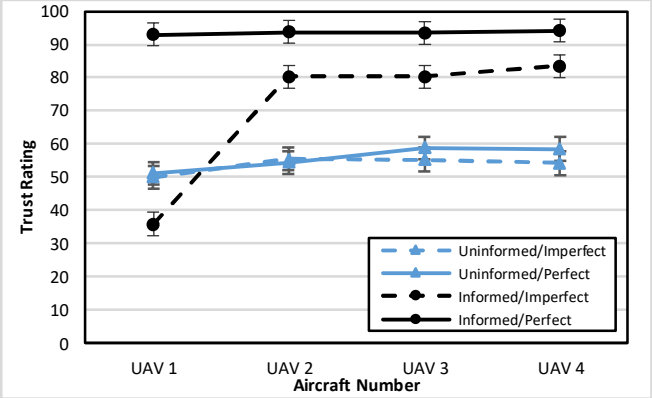

Figure 3. Post Task Subjective Trust Ratings with SEM error bars.

To sum up, these findings demonstrate that system-wide trust can occur in realistic multi-agent systems, and suggest that even perfect knowledge of system performance may not be sufficient to mitigate the effect. Complex interdependencies exist in multi-agent systems that can influence the ability to calibrate trust and appropriately rely on individual agents. Further system-wide trust research is warranted and should focus on methods to foster component specific trust as well as the limits of the effect. Also of interest is to explore the extent to which SWT strategy is employed when the system components are homogenous vs heterogeneous (i.e. four of the same UAVs vs four different UAVs). For example, a heterogeneous system may support CST, but it could be the case that CST is only the more accurate trust strategy if systems functions are indeed heterogeneous.

3. Effort #2: Examining risk—the effect of changing the stakes on operator trust and performance with autonomous systems

3.1 Motivation

The purpose of the 2nd effort detailed in this report was to examine the effect that a situational factor, risk, can have on an operator's willingness to trust in an autonomous teammate. Perceived risk is an important situational trust factor because an environment always involves some degree of uncertainty. Some authors even suggest that without some element of

risk, trust can be considered irrelevant (Wicks, Berman, & Jones, 1999). Examining risk in the laboratory presents a unique challenge because it is difficult to ensure that participants' have something at stake. Risk has, however, been extensively studied in the economics literature by manipulating the stakes involved in trust games. Trust games, such as the ultimatum game, require participants to trust a partner during a monetary exchange. It involves two participants, both of whom are given a sum of money at the outset. The first participant decides what portion of the money to send to the second participant, then the second participant simply accepts or rejects the proposed portion. If the second participant accepts, both participants are paid; if the second participant rejects, neither participant is paid (Guth, Schmittberger, & Schwarze, 1982). In other variations, the second participant may then either choose to keep the money passed or choose to send a portion back to the first participant. The amount of money passed by each participant is thought to reflect the trust the participant has in his or her partner (Berg, Dickhaut, & McCabe, 1995).

Despite the ultimatum game's widespread use to examine trust, it may not be the best tool to use in studies examining human-autonomy trust. For example, there has been some argument that the first participant's decision may reflect risk preferences, rather than trust, and instead researchers should solely examine the second participant's willingness to accept an offer, (Johannson-Stenman, Mahmud, and Marinsson, 2005). Studies employing this analysis have found that the second participant is more willing to accept a low offer with raised stakes (Cameron, 1999; Munier & Zaharia, 2002). Hence, while a useful laboratory paradigm, it may need to be amended when examining dynamic and cooperative teamwork. The aim of this second effort then was to investigate how differing levels of risk may impact trust in automation. Based on the results of Perkins et al. (2010) and studies using the ultimatum game, it was hypothesized that trust would decrease with increased risk. However, in contrast to past studies, this study 1) involves a tangible risk factor, 2) sought to utilize a behavioral measure of trust, and 3) involves a degree of interdependency/dependence on an autonomous partner to achieve a common goal. Past studies have often used subjective questionnaires to measure trust. These come with certain disadvantages. Subjective questionnaires only provide trust information after the experiment is finished and are prone to biases on the part of the participant. In this effort, a behavioral measure of trust was utilized that had been evaluated through extensive pilot testing in addition to subjective measures.

3.2 Experiment

Data was collected on thirty participants, and participants were randomly assigned to two levels of risk. The two levels of risk were defined by the amount of money a participant could potentially lose in the scenario by falling below a pre-defined performance criterion. All participants were given an amount of $50.00 before the scenario. Participants in the low risk condition stood to lose $10.00 for scoring below the performance criterion during the experiment, while participants in the high risk condition stood to lose $40.00. These parameters were based on a study conducted by Harinck et al. (2007) which found that participants will feel loss aversion if the payout is at least $40.00. Furthermore, criterion point values were established through extensive pilot testing. The Dynamic Distributed Decision Making (DDD) 4.0 simulation developed by Aptima Inc. was used in this experiment. DDD is a tool for creating human-in-the-loop distributed, multi-person and automation based scenarios. Participants performed the DDD scenario using a desktop computer and mouse. The scenario was a seven minute simulated counter air operation in which enemy targets entered and immediately began

moving towards a no-fly zone (red zone in this case). Participants controlled four UAV assets and an autonomous team member controlled four UAV assets. The participant and autonomous team member had separate, distinct zones for which they were responsible for protecting. The participant was responsible for protecting the yellow zone on the right and the autonomous team member was responsible for protecting the green zone on the left (Figure 4). This scenario has been executed successfully in previous studies of human-human teaming (e.g. Mckendrick et al., 2014).

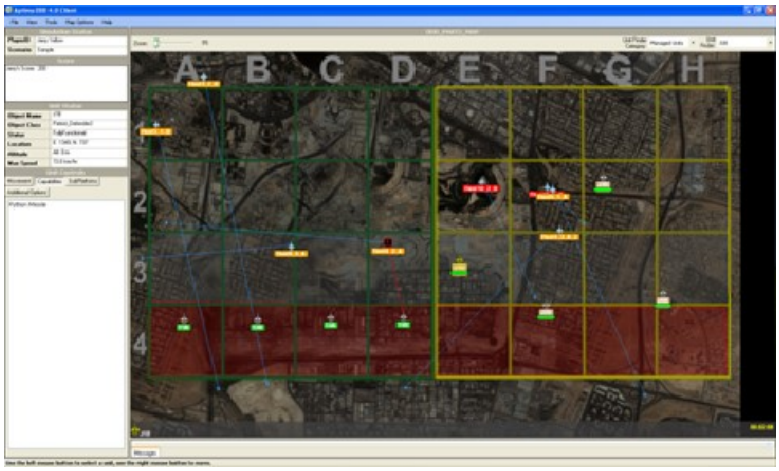

Figure 4. DDD simulation. Participant's AOR was the yellow grid. Automation's AOR was the green grid.

Participants were instructed that the goal of the scenario was to gain as many points as possible. A point system was explained as follows: 1) Participants lost 15 points for each orange enemy that entered the red zone. 2) Participants lost 30 points for each red enemy that entered the red zone. 3) Participants gained 50 points for destroying a red enemy while in the yellow zone. 4) Participants gained 50 points when the automation destroyed a red enemy in the green zone. 5) Participants gained 30 points for destroying a red enemy in the green zone. 6) Participants gained 25 points for destroying an orange enemy while in the yellow zone. 7) Participants gained 25 points when the automation destroyed an orange enemy in the green zone. 8) Participants gained 15 points for destroying an orange enemy in the green zone. This point system encouraged participants to make judgments about whether to intervene and destroy enemies in the autonomous team member's zone.

As in previous experiments that have used the DDD, percentage of enemy incursions into the red-zone was used to characterize operator performance, thus, lower numbers indicate better performance. The results of the experiment showed that there were more incursions in the Low Risk condition than in the High risk condition indicatignt that overall, performance was superior in situations of high risk. The results also showed that enemy incursions were higher for the autonomous teammate's area of responsibility (AOR) vs. the participant's AOR (Figure 5). This experiment also sought to validate a behavioral measure of trust—namely, the number of times the participants *intervened* in the autonomous teammates zone and adopted responsibility that was originally the job of the autonomous teammate. Thus, a high percentage of interventions are indicative of low trust. Results revealed that participants intervened imore in the high risk condition than in the low risk condition (Figure 6).

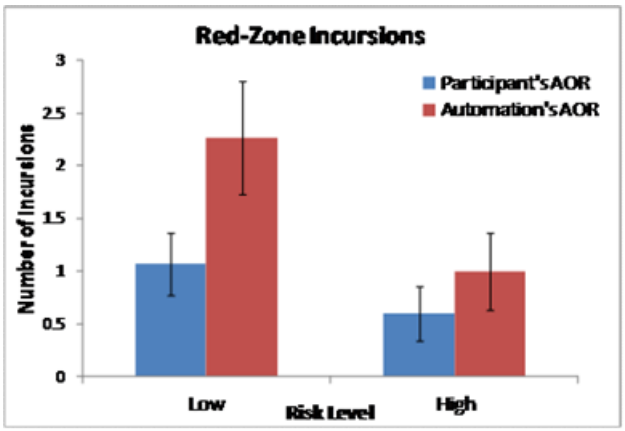

Figure 5. Number of enemies allowed into the red zone for both levels of risk and AOR. Error bars are standard error.

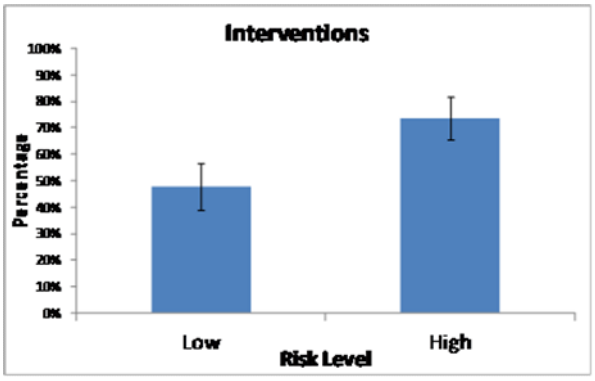

Figure 6. The percentage of interventions in the automation's AOR. Error bars are standard error.

To sum up, the results of this study have implications for how operators calibrate their trust in high stakes environments. While participants in the high risk condition intervened more, they did not score significantly higher than participants in the low risk condition. This pattern of results demonstrates that in a high risk situation, operators may under-rely on automation and unnecessarily increase their own workload. Overall, this study suggests that trust differs in situations of high risk and supports the need for developing behavioral measures of trust which are not prone to the biases associated with subjective measures. Future studies should examine how other factors that affect trust interact with risk, particularly the interaction of automation with extensive experience and also examining more moderate levels of risk.

4. Effort #3: Can autonomous systems be teammates?

4.1 Motivation

The purpose of the 3rd research effort was to explore the extent to which performance outcomes between humans and autonomous agents can be improved when the autonomous agent is treated as a team member as opposed to a tool. Technological advances have allowed us to overcome the limitations associated with "automation-as-a-tool" control strategies and shift towards the use of autonomous, self-governing systems. Unlike automated systems, which perform tasks typically performed by humans but still require human oversight, autonomous systems learn and are self-directed, eliminating the need for direct human control (de Visser, Pak, & Shaw, 2018; Hancock, 2017). While it is ideal to have a supervisory control framework for the governance of automated systems, the best way to structure groups of humans and autonomous systems are still being explored. Several possibilities exist, such as a hierarchical management structure, pure divisions of labor via task allocation, and specialized cliques (Groom & Nass, 2007). Yet another viable framework for human and autonomous agent collaboration is the more traditional team structure.

Teams are social structures characterized by high interdependence between team members and shared common goals (Salas, Cooke, & Rosen, 2008; Salas, Dickenson, Converse, and Tannenbaum, 1992). Team membership is associated with heightened communication, trust, effort, and commitment (Abrams, Wetherell, Cochrane, Hogg, & Turner, 1990). A comprehensive approach to evaluating human-autonomous agent teaming will consider not only team effectiveness (i.e. performance), but also critical team processes, such as the development of shared knowledge structures, emergent states, behavior patterns, and affect (Kozlowski & Ilgen, 2006). The most significant criticism lodged against the idea of human-autonomous agent teaming is that autonomous agents cannot perform essential teammate behaviors, such as the development of shared common goals, shared mental models, positive view of interdependence, fulfilled roles, and mutual trust (Groom & Nass, 2007). It should be noted, though, that these criticisms were raised at a time when the technology underlying autonomous systems had not been fully developed. Autonomous systems are now capable of sharing goals with a team and even disregarding human directions if they are perceived by the non-human agent as being counter to higher order goals (Schermerhorn & Scheutz, 2009). Others have explored augmented reality as a method to improve the sharing of mental models with autonomous systems (Green, Billinghurst, Chen & Chase, 2008; Michalos et al., 2015). These efforts suggest that the teamwork capabilities of autonomous systems will continue to expand, and exploring these capabilities in controlled research environments is a worthwhile endeavor.

Under this effort, two experiments were conducted to examine the extent to which social relations could improve teaming outcomes between autonomous systems and human agents. More specifically, using the comprehensive model of teamwork proposed above (cf. Kozlowski & Ilgen, 2006), we examined the extent to which team affect, team behavior processes, and team performance were improved by supporting the fundamental human need to interact with autonomous systems socially, as predicted by the CASA paradigm (Nass et al., 1994). In Experiment 1, we compared the social outcomes when autonomous agents were presented as teammates versus tools (i.e. a framing manipulation) and compared that directly to a condition where the human participant was paired with another human. In other words, we wanted to provide a comparison between the automation-as-a-tool and automation-as-a teammate

interaction paradigms. In Experiment 2, we evaluated the utility of employing a team building intervention, which have often been used in the human-human teaming literature, to examine the extent to which we could enhance social interactions between humans and non-human teammates. We predicted that if the team framing and team building intervention were effective, we would observe improved teaming outcomes at the affect, behavioral, and performance levels.

4.2 Experiment 1

Data was collected on sixty participants. Fifteen participants were randomly assigned to serve in one of four conditions defined by the factorial combination of a 2 × 2 between subjects design with agent type (human/autonomous) and organizational structure (tool/teammate) as the independent variables. In the autonomous agent condition, the second computer station was on and was visible to the participant, but there was no human or physical entity sitting at the station. In the human condition, a confederate sat at the second station. Those in the teamwork condition were informed that the experimental trial would be a multiplayer game in which the confederate was a teammate. Further guidance emphasized that participants should work to achieve the best team performance possible. Participants in the partner-as-a-tool condition were instructed to consider the confederate as a tool that could be directed or bypassed during gameplay.

The experimental platform used in the study was *Strike Group Defender* (developed by Metateq, Inc.), a serious game designed to train United States Navy personnel in ship defense techniques (see Figure 7). The game allowed for multiple players to participate in a single scenario, which created an environment that would benefit from teamwork. The participants were tasked to defend the strike group from incoming missiles while working with a partner that they believed to be either human or autonomous. Though effective teamwork was not essential to complete the task, better performance (i.e. a higher score) would result if participants employed teamwork behaviors such as coordination, communication, monitoring, and backup.

Teamwork was assessed by processes and outcomes in the same manner that human teams are typically evaluated (Kozlowski & Ilgen, 2006). Two processes were assessed during this study: Affect and Behavior. Affective processes were measured using the aforementioned subjective rating scales. Behavioral processes were assessed objectively by examining specific team behaviors such as adaptation, resource allocation, and communication behaviors. Several questionnaires were administered that assessed various team processes such as cohesion and trust.

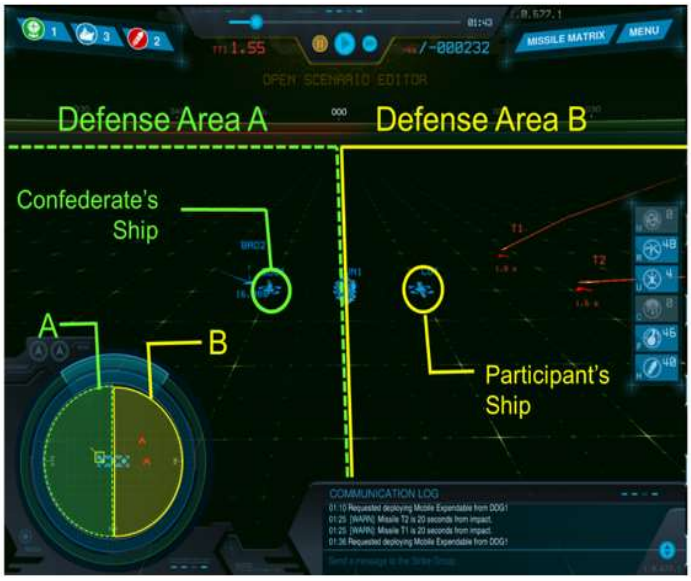

Figure 7. SGD interface showing Area Defense Strategy Division of Responsibility

An analysis was conducted to determine the effect of Agent Type (Human, Autonomous) and Structure (Team, Tool) on subjective measures of affect. In six out of ten measures, there was a significant main effect for Structure such that affect was higher when the interaction was structured as teamwork rather than a tool structure (Table 1). *Of note was the finding that there was no significant main effect or interaction for the performance measures.*

Table 1. Significance indicators for the various measures of affect (Exp 1)

	Structure	Agent Type	Interaction
Team Goals	↑	--	--
Perceived similarity	↑	--	--
Cohesion	↑	--	--
Trust	↑	--	--
Interdependence	↑	--	--
Confidence	↑	--	--
Collaborative climate	--	--	--
Team perception	--	--	--
Information quality	--	--	--
Role clarity	--	--	--

↑ = Statistically significant at .05 level (Teamwork > Tool)
-- = Not statistically significant

Experiment 1 was designed to compare social interactions between human and autonomous teammates in terms of affect, behavior, and performance outcomes when the interaction was framed as a team or tool. The findings of Experiment 1 related to affect demonstrate that framing the interaction with the autonomous agent as teamwork led to

improved subjective ratings of perceived similarity, interdependence, shared goals, confidence, cohesion, and trust, but no differences in team perception, information quality, role clarity, and collaborative climate. Perhaps the latter subjective measures did not statistically emerge in support of the teaming structure because participants did not have sufficient opportunity to share much task relevant information, discuss role assignments, or collaborate with the agent. Also, there was no difference in performance between any of the independent variables.

Taken together, the findings from Experiment 1 demonstrate that emphasizing a teamwork structure could potentially improve affective and behavioral outcomes between humans and autonomous teammates. Unfortunately, we cannot make those same assertions about performance outcomes since there were no differences in performance. This could potentially suggest that a team structure between human and autonomy is necessary but insufficient to produce desirable teamwork performance outcomes. In other words, while participants seemed to embrace the team structure with both humans and autonomous agents the framing was not enough to stimulate superior levels of interaction that could yield better performance results. *Thus, Experiment 2 was designed to delve more deeply into the idea that a teaming structure can be used to improve performance outcomes between humans and machines. More specifically, Experiment 2 was designed to see if we could employ a team building intervention that has been shown to successfully improve team performance outcomes amongst humans in the context of a human-machine partnership.*

4.3 Experiment 2

The purpose of Experiment 2 was to determine if improving the social interactions in human-autonomous agent teaming through team building interventions could improve team performance outcomes. Data was collected on sixty participants. The design of Experiment 2 was a 2 × 2 between subjects design with Agent Type (human/autonomous) and Team Building Type (informal/formal) as the independent variables. Similarly to Experiment 1, the Agent Type variable was defined by whether or not the participant's teammate was a human or autonomous agent. In the autonomous agent condition, the neighboring computer station would be active but there would be no human sitting at the station. In the human condition, a confederate would be sitting at the second station. The Team Building Type variable was defined by the way the participant and confederate interacted prior to beginning the missile defense scenario. In the informal team building condition, participants completed a non-task related cooperative game with the confederate. The participants played a freeware version of Tetris™ that has been adapted for team performance called Quadra. In the formal team building condition, participants engaged in a formal role clarification and goal setting exercise. Participants in the formal team building condition completed an online team building task comprised of goal setting and role clarification interventions. The purpose of this manipulation was to ensure that any observed changes in performance were attributable to the team building manipulation and were not a function of generalized cooperation or interaction. Other than the addition of the new manipulation, the procedure and methodology of experiment 2 was identical to that in experiment 1.

A 2 × 2 between-subjects ANOVA was conducted to determine the effect of Agent Type (Human, Autonomous) and Team Building Type (Informal, Formal) on subjective measures of affect. In nine out of ten measures, there was a significant main effect for Team Building Type such that affect was higher for participants in the formal team building condition relative to the informal condition.

Table 2. Significance indicators for the various measures of affect (Exp 2)

	Team Bldg Type	Agent Type	Interaction
Team Goals	↑↑	--	--
Perceived similarity	↑↑	--	↑
Cohesion	↑↑	--	--
Trust	↑	--	--
Interdependence	↑↑	--	--
Confidence	↑↑	--	--
Collaborative climate	--	--	--
Team perception	↑↑	--	--
Information quality	↑	--	--
Role clarity	↑	--	--

↑Significant at the .05 level (Formal Team Building > Informal Team Building).
↑↑Significant at the .01 level (Formal Team Building > Informal Team Building).

Unlike experiment 1, the analysis of performance in experiment 2 did reveal differences. There was a main effect for Team Building Type on all three scoring categories such that participants in the formal team building condition scored higher than those in the informal condition. Furthermore, there was a main effect for Agent Type on two of the three categories: overall score and results score. Participants in the autonomous agent condition outscored those in the human condition. The results of this analysis can be viewed in Table 3.

Table 3. Significance indicators for the various scoring categories (Exp 2).

	Team Building Type	Agent Type	Interaction
Overall Score	↑↑	↑↑	--
Results Score	↑	↑	--
Efficiency Score	↑↑	--	--

↑Significant at the .05 level (Formal Team Building / Human Agent > Informal Team Building / Autonomous Agent).
↑↑Significant at the .01 level (Formal Team Building / Human Agent > Informal Team Building / Autonomous Agent).

These findings demonstrate that formal team building interventions, which are designed to enhance social interactions, can improve teamwork outcomes relative to teams that do not receive such interventions. Predicted affect and performance outcomes were observed for nearly every measure. Furthermore, the effect of formal team building seemed to reduce differences between human and autonomous agent teammates that had been observed in Experiment 1. This study supports the need for consideration of social interactions between humans and autonomous agents. Beyond simply labeling an interaction as teamwork, it may be necessary to employ formal team building interventions, which are designed to improve social interactions.

References

Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology*, *29*(2), 97-119.

Beck, H. P., Dzindolet, M. T., & Pierce, L. G. (2007). Automation usage decisions: Controlling intent and appraisal errors in a target detection task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(3), 429-437.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122-142.

Cameron, L. A. (1999). Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry*, *37*(1), 47-59.

Defense Science Board Task Force Report: The Role of Autonomy in DoD Systems. *Department of Defense*, July 2012.

Dekker, S. W. A., & Woods, D. D. (2002).MABA-MABA or abracadabra?Progress on human automationcoordination.*Cognition, Technology, and Work, 4,* 240–244

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation'to 'autonomy': the importance of trust repair in human–machine interaction. *Ergonomics*, 1-19.

Geels-Blair, K., Rice, S., and Schwark, J. (2013). Using system-wide trust theory to reveal the contagion effects of automation false alarms and misses on compliance and reliance in a simulated aviation task. *The International Journal of Aviation Psychology, 22(3)*, 245-266.

Green, S. A., Billinghurst, M., Chen, X., & Chase, J. G. (2008). Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design. *International Journal of Advanced Robotic Systems*, *5*, 1-18.

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

Güth, W., Schmittberger, R., &Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, *3*(4), 367-388.

Hancock, P. A. (2017). Imposing limits on autonomous systems. *Ergonomics*, *60*, 284-291.

Harinck, F., Van Dijk, E., Van Beest, I., &Mersmann, P. (2007). When gains loom larger than losses reversed loss aversion for small amounts of money. *Psychological science*, *18*(12), 1099-1105.

Johansson-Stenman, O., Mahmud, M., & Martinsson, P. (2005). Does stake size matter in trust games? *Economics Letters*, *88*(3), 365-369.

Keller, D., & Rice, S. (2010). System-wide versus component-specific trust using multiple aids. The Journal of General Psychology, 137(1), 114-128.

Kozlowski, S. W., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, *7*, 77-124.

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

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

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*, 277-301.

Mckendrick, R., Shaw, T. H., Saqer, H., de Visser, E., Kidwell, B., &Parasuraman, R. (2014). Team performance in networked supervisory control of unmanned air vehicles: Effects of automation, working memory, and communication content. *Human Factors*, *56*, 463-475.

Michalos, G., Karagiannis, P., Makris, S., Tokçalar, Ö., & Chryssolouris, G. (2016). Augmented reality (AR) applications for supporting human-robot interactive cooperation. *Procedia CIRP*, *41*, 370-375.

Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science, 1,* 354–365.

Munier, B., &Zaharia, C. (2002). High stakes and acceptance behavior in ultimatum bargaining. *Theory and Decision*, *53*(3), 187-207.

Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78). ACM.

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

Rice, S., & Geels, K. (2010). Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *Journal of General Psychology*, *137*, 362–375.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*, 540-547.

Salas, E., Dickenson, T.L., Converse, S.A., & Tannenbaum, S.I. (1992). Toward an understanding of team performance and training. In R. Sweezy & E. Salas (Eds), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex.

Schermerhorn, P., & Scheutz, M. (2009, November). Dynamic robot autonomy: Investigating the effects of robot decision-making in a human-robot team task. In *Proceedings of the 2009 international conference on multimodal interfaces* (pp. 63-70). ACM.

Wicks, A. C., Berman, S. L., & Jones, T. M. (1999). The structure of optimal trust: Moral and strategic implications. *Academy of Management review, 24*, 99-116.

5. Training

Over the course of this three-year award, ten (10) Ph.D. students, and three (3) M.A. students, and one (1) B.S. student were trained in conjunction with this project.

Skills acquired through the project include teamwork, written and verbal communication, and dissemination of work through publications. This includes presentation skills (e.g., invited talks, poster presentations at a premier conference, and dissertation defenses). It should also be noted that a number of independent studies were supervised by the PI in related areas emphasizing design and implementation of experiments as well as writing skills.

This project contributes to human resource development by educating students through research, providing new educational materials in the classroom, and giving students the

communication and writing skills needed to advance science and engineering for future generations.

6. Publications

*Indicates graduate student author

*Walliser, J., de Visser, E.J., Wiese, E., Shaw, T.H. (2019). Team Structure and Team Building Improve Human-Machine Teaming with Autonomous Agents. *Journal of Cognitive Engineering and Decision Making.* doi.org/10.1177/1555343419867563

De Visser, E.J., Peeters, M. M .M., Jung, M.F., *Kohn, S., Shaw, T.H., Pak, R., Neerincx, M.A. (in press). Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams. *International Journal of Social Robotics.*

De Visser, E., Pak, R., Shaw, T.H. (2018). From "automation" to "autonomy": The importance of trust repair in human-machine interaction. *Ergonomics*, *61*, 1409-1427.

*Kohn, S. C., *Momen, A., Wiese, E., Lee, Y., and Shaw, T. H. (2019). The consequences of purposefulness and human-likeness on trust repair attempts made by self-driving vehicles. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 63, No. 1) Sage CA: Los Angeles, CA: SAGE Publications.

*Kohn, S. C., Quinn, D., Pak, R., de Visser, E. J., & Shaw, T. H. (2018). Trust repair strategies with self-driving vehicles: An exploratory study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62.

Kluck, M., *Kohn, S. C., Walliser, J. C., de Visser, E. J., Shaw, T. H. (2018). Stereotypical of us to stereotype them: The effect of system wide trust on heterogenous populations of unmanned autonomous vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 62.

*Walliser, J.C., *Mead, P.R., & Shaw, T.H. (2017).The perception of teamwork with an autonomous agent enhances affect and performance outcomes. *Proceedings of the Human Factors and Ergonomics Society, USA, 61.*

*Satterfield, K., Baldwin, C., de Visser, E., Shaw, T.H. (2017). The Influence of Risky Conditions on Trust in Autonomous Systems. *Proceedings of the Human Factors and Ergonomics Society, USA, 61.*

*Walliser, J.C., de Visser, E.J., Shaw, T.H. (2016). Application of a system-wide trust strategy when supervising multiple autonomous agents. *Proceedings of the Human Factors and Ergonomics Society, USA, 60.*

Walliser, J. C., Mead, P., & Shaw, T. H. (2017, June) Human-autonomous agent teaming: Improving teamwork outcomes with team building interventions. Paper presented at the 71st meeting of the Department of Defense Human Factors Engineering Technical Advisory Group, Atlantic City, NJ.

*Walliser, J.C., *Mead, P., & Shaw, T.H. (2016, May). *Can autonomous systems be teammates?* Paper presented at the 70th meeting of the Department of Defense Human Factors Engineering Technical Advisory Group, Hampton, VA.