Human Factors and Simulation, Vol. 83, 2023, 36–45
https://doi.org/10.54941/ahfe1003560

Measuring Trust in a Simulated Human Agent Team Task

Cherrise Ficke[1], Arianna Addis[2], Daniel Nguyen[3], Kendall Carmody[4], Amanda Thayer[5], Jessica Wildman[6], and Meredith Carroll[7]

[1]Advancing Technology-Interaction & Learning in Aviation Systems (ATLAS), Melbourne, FL 32091, USA
[2]Institutefor Culture, Collaboration and Management (ICCM), Melbourne, FL 32091, USA

ABSTRACT

Due to improvements in agent capabilities through technological advancements, the prevalence of human-agent teams (HATs) are expanding into more dynamic and complex environments. Prior research suggests that human trust in agents plays a pivotal role in the team's success and mission effectiveness (Yu et al., 2019; Kohn et al., 2020). Therefore, understanding and being able to accurately measure trust in HATs is critical. The literature presents numerous approaches to capture and measure trust in HATs, including behavioral indicators, self-report survey items, and physiological measures to capture and quantify trust. However, deciding when and which measures to use can be an overwhelming and tedious process. To combat this issue, we previously developed a theoretical framework to guide researchers in what measures to use and when to use them in a HAT context (Ficke et al., 2022). More specifically, we evaluated common measures of trust in HATs according to eight criteria and demonstrated the utility of different types of measures in various scenarios according to how dynamic trust was expected to be and how often teammates interacted with one another. In the current work, we operationalize this framework in a simulation-based research setting. In particular, we developed a simulated search and rescue task paradigm in which a human teammate interacts with two subteams of autonomous agents to identify and respond to targets such as enemies, improvised explosive devices (IEDs) and trapped civilians. Using the Ficke et al. (2022) framework as a guide, we identified self-report, behavioral, and physiological measures to capture human trust in their autonomous agent counterparts, at the individual, subteam, and full team levels. Measures included single-item and multi-item self report surveys, chosen due to their accessibility and prevalence across research domains, as well as their simplistic ability to assess multifaceted constructs. These self-report measures will also be used to assess convergent validity of newly developed unobtrusive (i.e., behavioral, physiological) measures of trust. Further, using the six-step Rational Approach to Developing Systems-based Measures (RADSM) process, we cross-referenced theory on trust with available data from the paradigm to develop context-appropriate behavioral measures of trust. The RADSM process differs from traditional data-led approaches in that it is simultaneously a top-down (data-driven) and bottom-up (theory-driven) approach (Orvis et al., 2013). Through this process, we identified a range of measures such as usage behaviors (to use or misuse an entity), monitoring behaviors, response time, and other context-specific actions to capture trust. We also incorporated tools to capture physiological responses, including electrocardiogram readings and galvanic skin responses. These measures will be utilized in a series of simulation-based experiments examining the effect of trust violation and repair strategies on trust as well as to evaluate the validity and reliability of the measurement framework. This paper will describe the methods used to identify, develop and/or implement these measures, the resulting measure implementation and how the resulting measurement toolbox maps onto the evaluation criteria (e.g., temporal resolution, diagnosticity), and guidance for implementation in other domains.

Keywords: Dynamic trust, Measurement framework, Autonomous agents, Unobtrusive measurements

INTRODUCTION

Recent advances in the technological domain have demonstrated that intelligent agents are able to accomplish a greater number of tasks of varying complexity, whilst working alongside human operators to successfully complete missions (Schaefer et al., 2017). For example, complex tasks such as suppression of enemy air defense and Intelligence Surveillance and Reconnaissance can be carried out by human agent teams (HATs). Consequently, it is expected the proliferation of automated agents will continue growing in prevalence (Hanscom & Bedford, 2013). As HATs operate in more complex tasks, appropriate calibration of a human's trust in agents is a pivotal factor that affects team dynamics and effectiveness (Kohn et al., 2020). In order to ensure effective HAT performance, operator trust in the agent must accurately reflect the agent's limitations and capabilities (Yu et al., 2019). The key to this is the ability to measure trust in order to understand how trust is developed and maintained.

The literature presents a range of measures to assess operator trust in autonomous agents that vary in effectiveness, depending on the measurement needs. Selecting the best measures to use in relation to the measurement goals can be a cumbersome process. To make this process easier and more straightforward, Ficke et al. (2022) developed a theoretical framework to guide researchers on the best trust measures to use for three potential use cases that varied in trust dynamics (i.e., how quickly trust is expected to change) and the levels of interactivity expected to occur between the human and agent. In the current effort, survey, behavioral, and physiological measures from the framework are being implemented in a simulated search and rescue context to capture human trust in a multi-agent team. The objective of this paper is to illustrate the operationalization of this trust measurement framework in a simulated HAT context.

METHODS

To measure trust in multi-agent HATs, the testbed for this study was designed to facilitate key elements of teamwork, such as coordination and backup behaviors, in order to study emergent states. As such, the testbed was designed to 1) support perceived agent autonomy, 2) facilitate task interdependence among team members, 3) enable multi-agent team compositions, and 4) incorporate missions long enough to introduce dynamic events and observe the proceeding effects over time (Nguyen et al., 2022). By leveraging these four principles, the resulting experimental task included an hour-long simulated search and rescue task in which participants follow an assigned route to search through houses to locate Improvised Explosive Devices (IED), survivors, and enemies. Two subgroups consist of one drone and one ground vehicle that work with the participant during the task and have unique capabilities. Specifically, ground vehicles are capable of clearing rubble blocking an entryway, and drones are capable of accessing rooftops to locate survivors or enemies. Due to the participants' inability to perform these actions themselves, it is essential they utilize their agent teammates to complete the mission successfully. This requires role interdependency between the agent teammates and the human operator, whilst providing a sufficient number of interactions to take place for trust development. To facilitate assessment of changes in trust, competence-based trust violations (i.e., a violation designed to influence an individual's perception of the agent's technical skills, experience, and abilities needed to carry out the assigned task), are injected during the simulation at various timings and frequencies. For example, participants are alerted that one of the agents failed to detect an IED, resulting in the death of surrounding civilians.

Measure Selection

To measure fluctuations in HAT trust in this performance context, multiple measures were selected. The selection of these measures was guided by the measurement framework illustrated in Ficke et al. (2022), in which a series of survey, behavioral, and physiological measures were evaluated against eight measurement criteria to assess the effectiveness of each measure. Evaluation criteria included: (a) sensitivity (i.e., how easily the measure detects levels of trust), (b) diagnosticity (i.e., how well the measure explains the "why" behind the trust levels), (c) selectivity (i.e., how well it distinguishes between trust and other constructs), (d) unobtrusiveness (i.e., how disruptive a measure is), (e) temporal resolution (i.e., how often data is recorded), (f) reliability (i.e., how consistent scores are), (g) affordability, and (h) resource intensiveness (i.e., time and energy needed to collect and analyze the data). Based on Ficke et al.'s use case with the highest degrees of anticipated trust fluctuations and interactivity, a suite of measures was selected. This included (a) a multi-item trust scale, chosen due to their high levels of sensitivity, reliability, and diagnosticity, (b) single-item survey measures, chosen due to their high levels of selectivity, (c) usage behaviors, chosen due to their high levels of unobtrusiveness, temporal resolution, and diagnosticity, (d) context-specific behaviors chosen due to their high levels of unobtrusiveness, temporal resolution, and diagnosticity, as well as physiological measures including (e) cardiovascular and (f) galvanic skin response (GSR) measures, chosen due to their high levels of temporal resolution.

Measure Review and Refinement

The following section describes our process and rationale for deciding which measures to employ in our experiment, given the myriad of options in the HAT literature. We begin by detailing how and why we chose certain survey measures, then outline the systems-based approach we utilized to design and implement behavioral and physiological measures of HAT trust.

TRUST MEASUREMENT SITE

The measure selection and refinement process led to the identification of six measures that were particularly suitable for our study. In the following section, we describe how these measures were specifically implemented within our simulation-based study.

Survey Measures

Many measures of trust exist for measuring trust across the HAT literature and organizational teams literature. Among the most popular of these are the Checklist for Trust between People and Automation (Jian et al., 2000) and Mayer & Davis's (1999) Measure of Trust and Trustworthiness. The Checklist for Trust between People and Automation (Jian et al., 2000) is a seminal scale for measuring trust in automation, and has pioneered the way for the factor structure of many scales to come after it based on the results of the factor analysis and cluster analysis conducted during its development. However, as noted by the authors, the results of these analyses were difficult to interpret in relation to scale dimensionality. Although trust-related words were grouped through them, the items were never officially labeled and grouped in order to capture specific dimensions of trust. Given this, we chose not to utilize this measure.

Meanwhile, in Mayer & Davis's (1999) study, specific dimensions of trust were parsed out into Trustworthiness components (Ability, Benevolence, Integrity), and Trust (Propensity to Trust and Trust). Results of their confirmatory factor analysis provided evidence that people can distinguish between trust itself and factors of trustworthiness. Although this measure is useful for this distinction and has since been commonly adapted into the HAT literature (Kohn et al., 2020), we decided not to use this measure for two reasons. First, our experimental focus did not necessitate the inclusion of the scale in its entirety (e.g., Benevolence and Integrity subscales). Second, we believe that propensity to trust is a theoretically distinct construct in that it is an individual difference which should be separated from trust as an attitude. As such, we wanted a measure that focused specifically on trust as an attitude, while measuring propensity to trust separately through its own validated scale.

In light of these scale properties and three specific needs for our research paradigm, we employed Wildman et al.'s (2009) Trust/Distrust Scale. First, we are specifically interested in competence-based trust out of concern for the saliency and ambivalence of an integrity-based violation in this study. Correspondingly, we required a trust scale that captured competence-based trust itself rather than as a factor of trust. Additionally, this scale is designed to be used with multiple referents, which is ideal given that we are measuring trust among individual team members and the team as a whole. Finally, trust is treated as a multidimensional construct.

In addition to the Wildman et al. (2009) scale, we used a single-item measure of trust written in-house which reads "How much do you trust each of the following to do what is expected of them during your mission from this point on?". Participants rate each agent and the team as whole on a 5-point scale ranging from "Not at all" to "A great deal". This item directly captures trust and also specifies the trustee (the drones and ground robots), which are the two important components of single item measures of trust (Körber, 2018). It also enables us to quickly measure trust at multiple points throughout the experiment, identify how trust changes dynamically over time, and offers a more detailed analysis of trust networks (Ficke et al., 2022).

It is worth noting that in addition to surveying trust itself, we measured well-known predictors of trust. Notably, we captured propensity to trust technology (Jessup et al., 2019) and personality (Donnellan et al., 2006). These were measured during the pre-study survey to collect demographic information, measure individual differences, and account for any impact they might have on trust regardless of the trust violations.

Self-Report Measures: Validated Scales and Single Items

Within our simulation-based experiment, we employ a validated trust scale as well as a single-item measure in order to obtain both validated forms of data and more frequent data directly from the participant. For the validated trust scale, we use a modified 8-item version of Wildman and colleagues' Trust & Distrust scale (2009). Although the original scale contains 16 items that target both competence-based and integrity-based trust, we did not incorporate integrity-based items since we focus on competence-based trust. The validated scale is given to participants before the simulation begins, 30 minutes into the middle of the simulation, and after the simulation ends. Although validated scales are highly selective, diagnostic, and reliable, and would ideally be used to capture trust whenever possible, this was the best option for employing the scales while minimizing disruptions to the participant's natural task flow within the simulation.

Given the dynamic nature of trust, we also employed a single item measure to capture trust more frequently. Every 10 minutes in the 60-minute simulation, a brief pop-up prompts participants to rate "how much they trust [each team member and the team as a whole] to do what is expected of them in the mission from this point on?", yielding a total of 6 measures of trust during the simulation. The timing of the trust violations vary across conditions, occurring at approximately 10, 30, or 50 minutes into the 60-minute simulation. The single item trust surveys are thus presented every 10 minutes in order to capture both the impact of these trust violations as well as consistently capturing trust dynamics throughout the entire simulation.

Behavioral and Physiological Measures

The Rational Approach to Developing Systems-based Measures (RADSM) was utilized to design and implement both behavioral and physiological measures of HAT trust. RADSM is a six-step process that takes both a top-down and bottom-up approach to measure development by using systems-based data to assess a variety of constructs, such as coordination or trust (Orvis et al., 2013). First, we identified the construct of interest, competence-based trust. Competence- based trust is derived from the trustor's belief of the trustee's skills, knowledge, and expertise to complete the task (Mayer et al., 1995). If an individual has competence-based trust in the context of our study, they will believe that the agent can adequately perform their role of searching for and identifying enemies and civilians, calling the human over for support, clearing rubble or scanning rooftops.

Second, we developed a list of observable behaviors that would demonstrate competence-based trust within our simulated search and rescue mission. We began this process by referring to a previously conducted literature review on trust dynamics in HATs. Out of the 167 papers in the literature review, we identified 38 that utilized unobtrusive measurement. From there, we coded each article for the data source and type, equipment or device used, analysis method, and any theoretical or empirical support for using the unobtrusive measure to capture the construct. After completing the coding process, we generated a list of ten observable behaviors demonstrating competence-based trust within our study. For example, someone with high trust in an agent might accept agent input more frequently (Wang et al., 2017) or monitor the team less. In contrast, humans with lower trust in an agent might frequently reject agent input or monitor their team more.

Third, we identified the type of data available and how that data will be analyzed. Four types of data are available in the experiment: audio, video, physiological, and testbed simulation data. Audio data is collected from verbal questions that researchers ask the participants at the midpoint and end of the experiment, and through any verbal expressions that the participant makes during the experiment. Video data is collected from a video of the participant and a screen capture of the participant completing the simulation. Physiological data is collected via the Shimmer3 GSR+ unit that captures participants' galvanic skin response (GSR) and electrocardiogram (ECG). Finally, data from the testbed captures information such as how many times the human calls the agent for help or how many times the human monitors the agents' progress via clicking on the summary statistics menu. After developing a list of general trust behaviors in Step 2, the team identified indicators specific to the study context that were aligned with each behavior. For example, the specific indicator aligned with team monitoring was the number of times that the participant clicked on the simulation's summary statistics menu to monitor agent (or team) progress. Finally, a list of data analysis methods (e.g., speech to text and keyword matching to analyze audio data) was also developed.

Fourth, the construct indicators developed in Step 2 were matched with the data sources and analyses identified in Step 3. In other words, this information is combined into an "item" and various "items" comprise our systems-based measure of competence-based trust. Orvis et al. (2013) remarks that this process is similar to creating survey-based measures, but systems-based measures may aggregate items differently based on the multilevel nature of some constructs. In this experiment, there are four measurement levels to consider: team, agent subteam, agent type, and individual. In the case of team monitoring for example, we can analyze how many times the human participant opens the summary statistics log and accesses statistics for a) the team as a whole, b) individual agents, c) each agent subteam, or d) each agent type (e.g., drones or ground vehicles).

Fifth, the measures will be instantiated. According to Orvis et al. (2013), instantiating the measures, or extracting the "items" developed in Step 4, is only possible after the data has been properly collected, stored, correlated and managed. Data analysis and validation have yet to occur, given that data collection is still in progress.

Sixth, once data is collected, construct validity will be demonstrated through three methods. First, subject matter experts will provide an assessment of face validity. Second, convergent validity will be examined by statistically testing whether the systems-based measures correlate with the eight competence-based trust and distrust items from Wildman et al. (2009) Trust/Distrust Scale and with a single item measure of trust developed by the research team. Third, discriminant validity will be assessed by examining the relationship between the systems-based measures with theoretically unrelated constructs (e.g., team cohesion, team effectiveness).

Behavioral Measures: Usage Behaviors & Context-Specific Behaviors

Within our simulation-based experiment, we capture usage behaviors based on how often the participant accepts an agent's request for assistance. Each time an agent requests the participant's assistance, the participant has the option to either accept or decline the request. Accepting the request is counted as using the agent, while ignoring the request is counted as discussing the agent. Frequencies of use behavior are thus positive reflections of trust (i.e., more use behaviors indicate higher trust levels), while disuse behaviors negatively reflect trust (i.e., more disuse behaviors indicate lower trust levels). Capturing these usage behaviors allows us to infer trust more frequently than is available through self-reported means.

To supplement usage behaviors and offer a more direct indicator of trust, multiple context-specific behaviors are captured within our simulation. First, a participant can check an agent's performance scores (e.g., how many targets they have addressed, how many mistakes they have made). This can be seen as monitoring behavior, and we infer that the quantity of this monitoring behavior is negatively related to their trust in the agent. Second, a person can send nudges to their agent teammates to request that a particular agent perform their tasks more quickly or more carefully. Similarly to monitoring behaviors, we infer that this nudging behavior is negatively related to their trust in the agent. Lastly, we record both the time it takes the human to respond to communications from their agent and the change in this response time as the mission progresses. When a person responds quickly, we infer they are more alert and diligent to respond to potential issues from the agent, and thus that faster response times are associated with less trust. Altogether, these context-specific behaviors contribute the additional sensitivity, diagnosticity, and selectivity for measuring trust that usage behaviors do not.

Physiological Measures: Electrocardiogram & Galvanic Skin Response

To capture more granular data and better identify when trust shifts, we capture two forms of physiological data. First, we capture cardiovascular data, which has been negatively correlated with trust (Tolston et al., 2018). More specifically, we employ an electrocardiogram (ECG), which measures the electrical activity of the heart in millivolts. Voltage itself can then be used to calculate heart rate, and subsequently, various heart rate states such as heart rate synchrony or variability, which have been negatively associated with trust (Tolston et al., 2018). Second, we employ a galvanic skin response (GSR) sensor, which measures skin conductivity in microsiemens. More specifically, microsiemens are useful for identifying the number of GSR peaks, which can be used to infer trust based on differences between a person's GSR peaks after a trust violation and their baseline (Khawaji et al., 2015). GSR activity has been found to be negatively correlated with trust (Hald et al., 2020).

It is important to note that in using physiological measures of trust that they serve as indicators of trust. Rather than being used as a direct measure of trust, ECG and GSR data act more like markers which may add additional insight to more diagnostic measures of trust such as context-specific behaviors and self-report measures. Whereas behavioral and self-report measures provide the selectivity, diagnosticity, and reliability to be confident that trust is indeed the construct being measured, physiological measures are intended to be paired with them in order to elucidate additional information about when and by how much trust may have shifted.

CONCLUSION AND FUTURE RESEARCH

This paper aimed to illustrate the development of HAT trust measures for a simulated search and rescue task based on the HAT trust measurement framework presented by Ficke et al. (2022), a framework developed to guide researchers down a more concise path when selecting measures to implement for a range of HAT use cases. Utilizing this framework, hand in hand with the RADSM process (Orvis et al., 2013), we developed a set of HAT trust measures for a simulated search and rescue task. This paper describes the process we used to identify and implement the measures and the resulting measures that will be utilized to capture and evaluate human trust in a multi-agent heterogeneous team.

Future research will include the collection of a large data set to facilitate evaluation of the validity of these trust measures. During data analysis, we will be examining ways to aggregate and combine the measures to further ensure the validity and the reliability of the framework.

As the current study aims to identify trust dynamics with competence-based trust violations, future research should also look to assess the reliability and validity of the framework in multi-agent HAT teams who are investigating integrity-based trust violations in a similar use-case. Integrity-based trust violations can be defined as an agent mis-prioritizing tasks, or having an integral misalignment with the goals of the human teammate, appearing to act on a set of unpredictable principles (Jensen et al., 2020). Future research should also look to investigate team trust dynamics when the team is introduced to various repair strategies. This would allow researchers to investigate a larger scope of trust dynamics and nuances among heterogeneous multi-HATs.

ACKNOWLEDGEMENT

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-21-1-0294. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Air Force.

REFERENCES

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychological Assessment*, *18*, 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Ficke, C., Carmody, K., Nguyen, D., Piasecki, I., Addis, A., Akib, M., Thayer, A. L., Wildman, J. L., & Carroll, M. (2022). Capturing trust dynamics in human-agent teams through unobtrusive measurement. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference (I/ITSEC)*, Orlando, FL.

Hanscom, A. F. B., & Bedford, M. A. (2013). Unmanned aircraft system (UAS) service demand 2015–2035. Literature review & projections of future usage. *Research and Innovative Technology Administration US Department of Transportation,* Washington, DC, USA.

Hald, K., Rehmn, M., & Moeslund, T. B. (2020). Human-Robot Trust Assessment Using Motion Tracking amp; Galvanic Skin Response. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 6282–6287. https://doi.org/10.1109/IROS45743.2020.9341267

Jensen, T., Khan, M. M. H., Albayram, Y., Fahim, M. A. A., Buck, R., & Coman, E. (2020). Anticipated Emotions in Initial Trust Evaluations of a Drone System Based on Performance and Process Information. *International Journal of Human–Computer Interaction*, *36*(4), 316–325. https://doi.org/10.1080/10447318.2019.1642616

Jessup, S., Schneider, T., Alarcon, G., Ryan, T., & Capiola, A. (2019). *The Measurement of the Propensity to Trust Automation* (pp. 476–489). https://doi.org/10.1007/978-3-030-21565-1_32

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. https://doi.org/10.1207/S15327566IJCE0401_04

Khawaji, A., Zhou, J., Chen, F., & Marcus, N. (2015). Using Galvanic Skin Response (GSR) to Measure Trust and Cognitive Load in the Text-Chat Environment. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1989–1994. https://doi.org/10.1145/2702613.2732766

Kohn, S. C., Kluck, M., & Shaw, T. (2020). *A Brief Review of Frequently Used Self-Report Measures of Trust in Automation*.

Körber, M. (2018). Theoretical considerations and development of a questionnaire to measure trust in automation. *Computer Security & Reliability*. Triennial Congress of the IEAF, Florence. https://doi.org/10.31234/osf.io/nfc45

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. https://doi.org/10.1037/0021-9010.84.1.123

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

Nguyen, D., Carmody, K., Wildman, J., & Carroll, M. (2022). Developing A Dynamic Testbed: Designing for Trust in Complex Human Agent Teams. *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, 1–4. https://doi.org/10.1109/ICHMS56717.2022.9980750

Orvis, K. L., DeCostanza, A. H., & Duchon, A. (2013). Developing systems-based performance measures: A rational approach. In *The interservice/industry training, simulation and education conference*. (I/ITSEC, no. 1).

Schaefer, K. E., Straub, E. R., Chen, J. Y. C., Putney, J., & Evans, A. W. (2017). Communicating intent to develop shared situation awareness and engender trust in human-agent teams. *Cognitive Systems Research*, *46*, 26–39. https://doi.org/10.1016/j.cogsys.2017.02.002

Tolston, M. T., Funke, G. J., Alarcon, G. M., Miller, B., Bowers, M. A., Gruenwald, C., & Capiola, A. (2018). Have a Heart: Predictability of Trust in an Autonomous Agent Teammate through Team-Level Measures of Heart Rate Synchrony and Arousal. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*, 714–715. https://doi.org/10.1177/1541931218621162

Wang, N., Pynadath, D. V., Hill, S. G., & Merchant, C. (2017). The Dynamics of Human-Agent Trust with POMDP-Generated Explanations. In J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, & S. Kopp (Eds.), *Intelligent Virtual Agents* (Vol. 10498, pp. 459–462). Springer International Publishing. https://doi.org/10.1007/978-3-319-67401-8_58

Wildman, J., Fiore, S., & Salas, E. (2009, July). Toward measurement of trust and distrust: Separate constructs or ends of a continuum? Poster session presented at the annual meeting of the *Interdisciplinary Network for Group Research*, Colorado Springs, CO.

Yu, K., Berkovsky, S., Taib, R., Zhou, J., & Chen, F. (2019). Do I trust my machine teammate? An investigation from perception to decision. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 460–468. https://doi.org/10.1145/3301275.3302277