

Agent Transparency and Reliability in Human–Robot Interaction: The Influence on User Confidence and Perceived Reliability

Julia L. Wright , *Member, IEEE*, Jessie Y. C. Chen , *Member, IEEE*, and Shan G. Lakhmani 

Abstract—Agent transparency is an important contributor to human performance, situation awareness (SA), and trust in human–agent teaming. However, agent transparency’s effects on human performance when the agent is unreliable have yet to be examined. This paper examined how the transparency and reliability of an autonomous robotic squad member (ASM) affected a human observer’s task performance, workload, SA, trust in the robot, and perceptions of the robot. In a 2 (ASM transparency) \times 2 (ASM reliability) within-subject design experiment, participants monitored a simulated soldier squad that included an ASM as it traversed a simulated training environment, while concurrently monitoring the environment for targets. There was no difference in participants’ performance on the target detection task, workload, or SA due to either ASM transparency or reliability. ASM reliability influenced participant trust and perceptions of the robot. Results suggest that reliability may be a stronger influence on the human’s perceptions of the robot than transparency. Robot errors had a profound and lasting effect on the participants’ perception of the robot’s future reliability and resulted in reduced confidence in their assessments of the robot’s reliability. These findings could have important implications for the continued use of automated systems when the user is aware of system errors.

Index Terms—Agent transparency, confidence, human–robot interaction, reliability, situation awareness (SA), trust.

I. INTRODUCTION

DEVELOPMENT of autonomous robotic agents for use in military operations is a priority for the U.S. military [1]. As a robotic agent’s autonomy increases, so too does the difficulty its human teammates’ experience in maintaining their awareness and understanding of the robot’s actions. Making the robot convey information to the human that would support a transparent human–robot interaction addresses these issues. In the context of human interaction with automated systems,

an approach to transparency centering on error and system reliability has been considered sufficient, however human–agent interaction requires a more complex approach to transparency [2]. Agent transparency has been described as supporting an operator’s comprehension about an intelligent agent’s intent, performance, future plans, and reasoning process [3]. To achieve transparency, the agent can use its interface elements—visual and other modalities—to convey information about itself to the human. Prior studies have explored aspects of transparency related to interface design, layout, and graphics, and found that relatively simple graphic designs can be used to convey large amounts of information with little to no increase in perceived workload [4]–[6]. A more complex issue is identifying what information should be conveyed, and then how much.

The situation awareness-based agent transparency (SAT) model [3], [7] was developed to provide a framework as to what information should be conveyed by an agent and how that information should be structured to support the human’s situation awareness (SA). By making the underlying processes that the autonomous agent uses to make its decisions, actions, and projections available to the human, the agent facilitates its operator’s SA of both the environment and itself, keeping the human in the loop [7]–[9]. Agent transparency has been identified as an essential factor in calibrating operator trust in an agent [2], [10], as well as being critical for the development of appropriate SA in human–robot teams [11], [12].

The SAT model is comprised of three levels of information, detailing the agent’s understanding of both the environment and itself [3], [7]. Level 1 is the basic information about the autonomous agent’s actions and plans, encompassing the agent’s knowledge of the environment and events within it. Level 2 is the agent’s reasoning process behind its actions/decisions, e.g., current priorities or constraints and affordances that the agent considered. Level 3 is the agent’s projected outcomes of its current plans/actions, predictions about its future actions, and state of uncertainty. Researchers have used the SAT model to inform human–robot interfaces in different agent paradigms, ranging from virtual decision aides to physically instantiated robotic teammates [6], [13], [14].

A. Transparency and Reliability

The autonomous squad member (ASM) project explored human interaction issues between a human and a robotic team member within a simulated military environment [1]. The ASM

Manuscript received October 9, 2018; revised February 12, 2019 and May 30, 2019; accepted June 11, 2019. This work was supported in part by the U.S. Department of Defense Autonomy Research Pilot Initiative (Program Manager Dr. D. H. Kim) and in part by the US Army Research Laboratory - Human-Robot Interaction program (Manager Dr. Susan Hill). This paper was recommended by Associate Editor J. Lyons. (*Corresponding author: Julia L. Wright.*)

J. L. Wright and S. G. Lakhmani are with the U.S. Army Research Laboratory, Human Research and Engineering Directorate, Orlando, FL 32826 USA (e-mail: Julia.l.wright8.civ@mail.mil; shan.g.lakhmani.civ@mail.mil).

J. Y. C. Chen is with the U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, MD 21005-5425 USA (e-mail: yun-sheng.c.chen.civ@mail.mil).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/THMS.2019.2925717

is a robotic mule that accompanies a dismounted soldier squad. The ASM's behavior and decision-making process are based on the goal-driven autonomy goal-reasoning model [15]. Under this model, agents use in-depth information from itself and its environment to select from a list of goals, priorities, and future expectations. Previous ASM studies have explored how this surface-level decision-making information could be made available to the human and how seeing these different configurations of information affected the human's awareness of the ASM's actions, reasoning, perceptions, and projected outcomes, as well as the human's perceptions of the ASM itself [4], [5]. Utilizing an icon-based (at-a-glance) display [3], [16], the ASM shared information about its current goal (SAT1), current priority (SAT2), and projected resource expenditure (SAT3). For example, the ASM may indicate that its current goal is to return to base, its current priority is to save time, and its expected resource expenditure is that it will use extra fuel to meet this goal with this priority. In this manner, the ASM can convey its current state. However, in those prior studies, the ASM did not reveal why it chose its current goal or priority, i.e., it did not share in-depth reasoning information.

In-depth SAT information describes the underlying factors that agents use to determine their goals, priorities, and projections. For instance, if the surface SAT level 2 information is that the robot has chosen a plan because it is trying to preserve its mechanical integrity, the in-depth information that motivated the robot's prioritization may be its observation of explosions or gunfire in the area. This in-depth information can also be used to help users determine why a robot made an error. If in the instance mentioned above the robot prioritized its mechanical integrity in error, the human could see that it did so because of spurious observations of explosions or gunfire.

Reliability is a factor that influences the proper use of automation in multiple ways. Unreliable automation has been found to have a negative impact on a user's task performance, trust in the system, and can result in disuse of the automation [18]–[21]. On the other hand, there are also examples where highly reliable automation was detrimental to human performance, resulting in overtrust and complacent behavior [7], [13], [19], [22]. Clearly, the relationship between agent reliability, transparency, and human performance is complex. Agent transparency could have a mitigating influence on the effects of agent reliability. However, agent transparency's impact on human performance when automation is unreliable has yet to be explored. In this paper, ASM reliability was manipulated to examine how it interacted with ASM transparency.

B. Human Factors in Human–Robot Teams

In human–agent teams, the human's perceived workload, their trust in the agent, and their perceived humanness of the agent have been shown to influence performance outcomes [5], [6], [8], [9], [13], [21], as such each of these factors will also be examined. Multiple factors inform an individual's decision-making process, such as their individual SA or meta-awareness factors (e.g., their confidence in their SA) [23], [24].

1) *Situation Awareness*: Developing appropriate SA is a mission-critical goal for human–robot teams [11]. SA is defined as “the perception of the elements in the environments within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” [9]. As such, SA refers to an individual's dynamic understanding of “what is going on” in a given system [25]. Forming and maintaining SA is an iterative process, and the information the person receives influences how the person accepts and incorporates new information into their mental model [23]. This anticipatory thinking heavily influences their decision-making process, as does meta-awareness factors, such as their confidence in their SA [23], [24]. As the SAT model was designed to support the operator's SA, the operator's SA was assessed to evaluate the effectiveness of the SAT information. Confidence in SA has also been identified as a crucial element in effective decision-making and performance [26], [27]. Operators' confidence in each of their levels of SA was assessed to determine their eventual willingness to act upon the received SAT information [24].

2) *Cognitive Workload*: The addition of SAT information may increase visual complexity since each increase in information adds more visual elements to the interface. As visual complexity increases, mental workload also tends to increase [28], [29]. Counterintuitively, previous studies that added visual interface elements to support different SAT levels did not report additional mental workload [6], [22]. This paper assesses how the increase in SAT information interacts with ASM reliability, and if that interaction results in a concomitant increase in workload.

3) *Trust*: Operator trust has been defined as: “the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability” [7]. Appropriate trust is essential to effective human–agent teaming; too much trust can result in operator complacency or misuse, while too little trust can result in disuse [21]. To provide support for the operator's trust calibration, a robot needs to provide meaningful insight into its actions and why it is performing them [3], [7]. Understanding the rationale behind an agent's behavior is crucial to the human's development and maintenance of appropriate trust and reliance upon the agent [2], [9], [17]. Information sharing has been shown to bolster the trust and liking of team members in human–human teams [30]. This study examines how the participant's trust in the agent was affected by errors of commission (correct actions in inappropriate contexts) that the agent committed [31], and if the display of information to support transparency information aided in mitigating the impact of the errors [32].

4) *Humanness*: Humanness includes both characteristics that are uniquely human (e.g., morality, civility, etc.) and those that are part of human nature (e.g., cognitive flexibility, emotionality, etc.) [33]. Robots that share information regarding their uncertainty and projections are viewed as being more like a human, compared to robots that only share their reasoning and current understanding of the environment [14]. The humanness of the robot is important in how the human perceives the robot, which in turn can affect their trust in the robot, and, if violated, how that trust could effectively be repaired [34]. In light of

this finding, this study examines how the robots' transparency and reliability affect the human's attributions of humanness and intelligence in the robot.

C. Current Study

This paper investigated how agent reliability and agent transparency, within the context of a supervisory control task, influence human behavior and attributions regarding the robot in a simulated military environment. The findings of this study are expected to show how the amount of agent transparency available to the human interacts with the agent's reliability to influence human performance, perceived workload, and perceptions of the robot.

The participant's role was to monitor a simulated dismounted soldier team, accompanied by an ASM, as it traversed a training course. Participants were given two sets of tasks. First, the participant was to monitor and evaluate the ASM, specifically its ability to correctly identify and respond to the events as each occurred. Second, participants were asked to detect threats in the surrounding environments and identify events the squad encountered. Participants completed four trials, each with a different ASM in terms of robot transparency and reliability. *ASM Transparency* and *ASM Reliability* were manipulated in each of the trials. Within-subjects evaluations, across transparency and reliability levels, compared differences in human performance on the target detection and event identification tasks, workload, SA, trust, and attributions regarding the robot. The following hypotheses are addressed.

- 1) Access to in-depth SAT information will be detrimental to participants' target detection performance (H1), or mental workload (H3) when the agent is unreliable, but will not affect participant's target detection performance, or mental workload when the agent is perfectly reliable.
- 2) Reduced agent reliability will reduce the participant's target detection performance (H2) and increase operator mental workload (H4), regardless of information level.
- 3) Access to in-depth agent SAT information will increase operator's SA of the agent (H5), trust in the agent (H9), and perceived humanness and intelligence of the agent (H11) when the agent is perfectly reliable, but will not increase operator's SA of the agent, trust in the agent, or perceived humanness and intelligence of the agent when the agent is less reliable.
- 4) Reduced agent reliability will decrease operator's SA of the agent (H6), trust in the agent (H10), and perceived humanness and intelligence of the agent (H12), regardless of information level.
- 5) Access to in-depth agent SAT information will increase operator confidence in their SA assessments (H7), regardless of agent reliability. Reduced agent reliability will decrease operator confidence in their SA assessments (H8), regardless of information level.

II. METHOD

A. Participants

A total of 56 participants (32 males, 23 females, and 1 unreported; $\text{Min}_{\text{age}} = 18$ years, $\text{Max}_{\text{age}} = 31$ years, and $M_{\text{age}} =$



Fig. 1. ASM experimental interface. The left-side monitor displays the lead soldier's point of view of the task environment. The right-side monitor displays the ASM's communication interface.

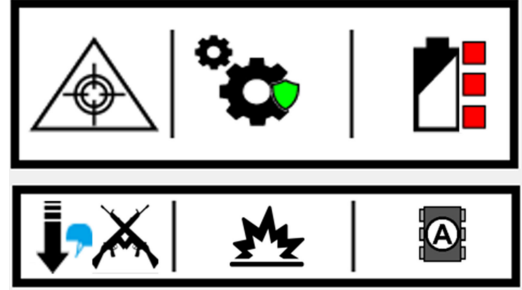


Fig. 2. ASM at-a-glance module. Surface-level (S) information is displayed in the upper frame. This image denotes the ASM has determined that there is a shooter (LH frame), it should prioritize preserving its mechanical integrity (center frame), and doing so will result in high energy expenditure (RH frame). In-depth (D) information is displayed by adding the lower module. This image indicates that the ASM has determined that there is a shooter present because it sees the squad getting down and returning fire; the ASM needs to act to preserve mechanical integrity because it has detected loud gunfire; the ASM will have high energy expenditure because it will take avoidance measures.

20.5 years) successfully completed the experiment. Participants received a cash payment (\$15/h) as compensation. The study took ~ 3 h to complete.

B. Simulator

A custom software application was used to present the ASM display to the participant. The simulation was delivered via a commercial desktop computer system, two 22" monitors, standard keyboard, and three-button mouse (see Fig. 1).

C. Experiment Design and Performance Measures

1) *Design*: The study was a 2 (ASM Transparency) \times 2 (ASM Reliability) within-subjects design experiment.

ASM Transparency was manipulated by varying the depth of SAT model information displayed (see Fig. 2). In the surface-level SAT model information (S) condition, the interface displayed the at-a-glance module with information pertaining to all three SAT levels (SAT 1, SAT2, and SAT3), similar to the one used by Selkowitz *et al.* [5]. The ASM identified what event was occurring (LH frame), what action it should take in response to the event (center frame), and the expected outcome of this action (RH frame). In the in-depth SAT model information (D) condition, a secondary module was added below the at-a-glance module, which depicted the underlying factors that led to each specific "decision" mentioned above.

		ASM Transparency Level	
		Surface level information only (S)	Surface level plus in-depth information (D)
ASM Reliability	100% Reliable (R)	100% Reliable/Surface level information (SR)	100% Reliable/Surface level plus in-depth info (DR)
	67% Reliable (U)	67% Reliable/Surface level information (SU)	67% Reliable/Surface level plus in-depth info (DU)

Fig. 3. ASM reliability by ASM transparency level design matrix.

ASM Reliability was manipulated by varying the ASM's error rate when it responded to events occurring in the environment. The simulated ASM determined what event was occurring by observing the squad's actions, and its subsequent behaviors were based on its identification of the event. Errors would occur when the ASM misinterpreted the squad actions, thus misidentifying the event, and responded as if a different event was occurring. An example of such an error would be the squad encounters a potential IED and seeks cover; the ASM misinterprets their behavior as responding to an ambush and deploys smoke grenades. The ASM was either 67% reliable (U) or 100% reliable (R), similar to the rates used by Mercado *et al.* [6]. In a 2×2 , counterbalanced, within-subjects design experiment each participant completed four trials, one in each of the conditions (i.e., SR, SU, DR, DU, etc.), see Fig. 3.

2) *Target Detection Task*: During each mission, participants conducted a target detection task wherein they were assigned to click on a specific vehicle whenever it appeared in the left-hand simulation screen. Each mission contained 15 targets and 80 "noise" vehicles. The participant encountered the vehicles at a rate of approximately eight vehicles per minute. To avoid potential ceiling or floor effects, participants' efficiency at this task was assessed. First, the correct targets score was calculated by finding the ratio of correct detections to total targets ($\#correct/\#total$ targets). Then, the efficiency score was calculated by finding the ratio of correct detections to the total number of clicks ($\#correct/\#total$ clicks). Finally, the target detection correctness efficiency (TDCE) score was calculated by multiplying the correct targets score by the efficiency score (correct targets \times efficiency).

3) *Event Identification Task*: The soldier squad encountered six events in each scenario (i.e., ambush, civilians, flooding, IED, obstacle, and sniper). Event task (ET) score reflects participants' average time to identify events and assesses correctness by including a time penalty for each incorrect response. The time penalty is the overall average response time from all conditions (i.e., 26.35 s). Lower ET scores indicate better event identification performance.

4) *Workload*: Participants' perceived workload was evaluated after each mission using the computerized version of the NASA-TLX [35].

5) *SA Scores*: To assess the participant's current awareness of their environment and the ASM, SA global assessment technique style SA queries were employed [36], [37]. During each

1. The system is deceptive when...

	hot at all			neutral		extremely	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

Fig. 4. Example question from the functional trust scale.

mission, the simulation was paused after each event, and the participant answered queries designed to assess their SA of the ASM and their confidence in their SA. SA1 and SA2 queries were scored as correct (+1) or incorrect (−1), with higher scores indicating better SA. SA3, however, was handled differently.

Typically, SA3 is assessed through queries about projected outcomes that can be predicted using SA1 and SA2 information available in the environment. Due to the nature of the task environment (episodic, stochastic, etc.) [38], the information available during the event could not help either the ASM or human predict the next event. However, participants could use the information available to establish expectations as to the ASM's future reliability, which could influence the participant's mental model and ongoing SA of the robot [39]. For the SA3 queries, participants were asked how reliable the ASM would be in the following event, based upon its performance in the current event. For purposes of this query, reliability was described as how well the robot correctly identified and responded to the event the squad encountered. Projected reliability (PR) was scored using a 4-point Likert scale (e.g., 4 = reliable, 1 = unreliable, etc.), with higher numbers indicating greater reliability. As such, if the ASM displayed reliable behavior in the current event, the projections for the following event should be "reliable" (score 4/4).

In addition to SA, the related concept of confidence in one's SA was assessed [26], [27]. The participant rated their level of confidence in their response to each SA query using 5-point Likert scales (e.g., 5 = very confident, 1 = not confident, etc.). Higher values indicate greater participant confidence.

6) *Trust*: After each mission, the functional trust survey was administered to assess the participants' trust in the ASM (see Fig. 4). The functional trust survey was developed in-house to distinguish the basis of an individual's trust in an autonomous agent. It is comprised of Jian *et al.*'s trust in automation survey [40], which had been modified to assess trust along the four functions of automation use (i.e., A: gathering and filtering information, B: integrating and displaying information, C: suggesting or making decisions, and D: executing actions), which yields a total of 64 survey items [41]. The length and complexity of the survey have the potential to induce respondent fatigue in participants, so previous research compared the methods of possibly mitigating these factors [42]. One study successfully used this survey to measure participants' state of trust after a single interaction [43]. In a follow-up study, they also successfully used this survey through multiple interactions to track the development of trust.

7) *Humanness*: The godspeed questionnaire series (GQS) was administered after each mission [44]. The GQS assesses an individual's perceptions of a robot on the attributes anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety via a series of bipolar Likert scale evaluations.

D. Procedure

After being briefed on the purpose of the study and signing the informed consent form, participants completed a demographics survey and began training. Training was self-paced and took 60–90 min to complete. Participants were told that the ASM was a robot that uses sensors to collect and analyze information about its environment, including the actions of the soldier teammates, and then uses that information when making decisions and executing actions. They were also shown an image of the ASM, which was simulated as a six-wheeled robotic cart with no visible appendages. Participants received training on the elements of the ASM interface, their specific tasks, and how to evaluate the ASM. The training included quizzes to assess their understanding and retention of the training content, and those who scored too low on the quizzes were allowed to review the information again and be reassessed. After completing the training, there was a practice session to familiarize the participants with the experimental environment. Participants were allowed a short (under 5 min) break before continuing to the experiment.

During the experiment, participants monitored a simulated dismounted soldier team, accompanied by an ASM, as it traversed a training course. In each trial, the soldier team took a different route through the course and was accompanied by a different ASM. As the team traversed the training course, events would occur, and the team (and ASM) would have to respond accordingly. Participants were required to identify the event, as well as detect other potential threats. After each event, the simulation would freeze, and the participants received SA queries related to the event. After answering the SA queries, the simulation would resume.

Each mission simulation lasted 12–15 min. After each mission, participants assessed their perceived workload, trust in the ASM, and humanness perceptions of the ASM. Participants completed four trials (a mission in each condition, followed by the related surveys) without additional rest breaks. This portion of the study took about 90 min to complete. Upon completion, participants were debriefed and dismissed.

III. RESULTS

SPSS V24 software was used to analyze the data. Data were examined using repeated-measures ANOVAs ($\alpha = 0.05$), and the Greenhouse-Geisser correction for sphericity is reported when applicable. Planned comparisons were conducted using paired t-tests to examine differences between conditions, specifically, SR–SU, SR–DR, SU–DU, and DR–DU. Unless otherwise specified, ANOVA results are reported in Table I, and planned comparisons are reported in Table II.

A. Task Performance

Participant performance on the target detection task was evaluated using the target detection correctness efficiency (TDCE) score, and performance on the event identification task was evaluated using the ET score. No difference in TDCE or ET scores due to information level or ASM reliability was found. Although the target detection and event identification tasks were

TABLE I
REPEATED-MEASURES ANOVA RESULTS

Metric	Measure	df_1	df_2	F	p	ω^2
Task Perf.	TDCE	3.00	53.00	0.12	.945	.00
	ET	3.00	53.00	0.88	.457	.00
Workload	NASA-TLX	3.00	53.00	0.52	.671	.00
SA1	% Correct	2.65	148.65	0.39	.733	.01
	Confidence	3.00	168.00	1.23	.300	.00
SA2	% Correct	3.00	168.00	1.57	.199	.01
	Confidence	3.00	168.00	0.81	.492	.00
PR	Score	2.39	131.49	114.29	<.001	.65
	Confidence	2.33	128.38	13.31	<.001	.56
Functional Trust Survey	Overall	3.00	52.00	4.41	.008	.15
	A	3.00	52.00	3.91	.014	.13
	B	3.00	52.00	1.92	.138	.05
	C	3.00	52.00	5.21	.003	.18
	D	3.00	52.00	5.98	.001	.21
Godspeed	Anthro.	2.55	140.28	16.42	<.001	.15
	Animacy	2.41	132.69	20.91	<.001	.21
	Likeability	2.39	131.42	26.68	<.001	.27
	Perc. Intell.	2.27	124.77	62.75	<.001	.50
	Perc. Safety	2.58	141.93	11.30	<.001	.13

not challenging and the event rate was quite low, it was still expected that the unreliable conditions would hamper performance (H1) due to increased cognitive load, and in-depth information presented during the unreliable condition would do so even more (H2). These hypotheses were not supported; there was no difference in performance regardless of agent reliability or transparency level.

B. Workload

Participant perceived cognitive workload was assessed using the NASA-TLX survey. Perceived workload was expected to increase in the unreliable conditions (H3) [21], and the addition of in-depth information in the unreliable condition would further increase workload (H4), but these hypotheses were not supported. There was no significant difference in global NASA-TLX scores due to either information level or ASM reliability.

C. Situation Awareness

There was no significant difference in the percentage of correct answers or participant confidence in their SA1 or SA2 responses between the experimental conditions (see Table I). Overall, the percent correct was over 83% for both SA1 and SA2 queries, regardless of condition, with confidence ratings of 4.6/5 and higher. As such, participants' SA1 (perception) and SA2 (comprehension) scores were uniformly high throughout the scenarios, which did not support the expectation (i.e., H5 and H6) that SA would be positively affected when the agent was perfectly reliable. Contrary to expectations (i.e., H7 and H8), there was no difference in the self-confidence ratings of SA1 and SA2 responses.

There were significant differences in participant projections of the ASM's reliability (PR) and their confidence in their evaluations between the experimental conditions (see Table I). Paired t-tests indicated that these differences were due to ASM reliability rather than ASM transparency (see Table II), thus not

TABLE II
BETWEEN-CONDITION PLANNED COMPARISONS RESULTS

Measure	Planned Comp.	ΔM	SE	p	d_s	% Diff.
PR Rating	SR > SU	0.79	0.06	<.001	2.46	20.53
	SR \approx DR	0.02	0.04	.524	0.10	0.61
	SU \approx DU	-0.01	0.07	.924	-0.02	-0.20
	DR > DU	0.76	0.06	<.001	2.21	19.88
Conf. in PR Rating	SR > SU	0.28	0.07	<.001	0.53	6.09
	SR \approx DR	-0.02	0.06	.755	-0.04	-0.38
	SU \approx DU	-0.05	0.04	.264	-0.08	-1.04
	DR > DU	0.25	0.06	<.001	0.53	5.47
Overall Trust	SR > SU	12.21	5.29	.032	0.90	18.25
	SR \approx DR	-2.19	3.98	.586	-0.21	-3.28
	SU \approx DU	-3.71	5.28	.491	-0.27	-6.77
	DR > DU	10.70	3.96	.012	0.99	15.48
Auto. Function A	SR > SU	11.92	5.68	.048	0.82	17.53
	SR \approx DR	-3.13	3.93	.434	-0.31	-4.61
	SU \approx DU	-7.39	5.46	.192	-0.53	-13.18
	DR > DU	7.67	3.61	.042	0.78	10.78
Auto. Function B	SR > SU	8.31	5.11	.110	0.58	12.12
	SR \approx DR	-1.33	4.94	.789	-0.13	-1.94
	SU \approx DU	-1.44	4.94	.772	-0.09	-2.38
	DR > DU	8.20	4.76	.091	0.70	11.74
Auto. Function C	SR > SU	14.00	5.54	.020	0.99	21.64
	SR \approx DR	-2.91	4.41	.515	-0.25	-4.49
	SU \approx DU	-3.84	5.61	.501	-0.26	-7.58
	DR > DU	13.07	4.50	.007	1.06	19.33
Auto. Function D	SR > SU	14.62	5.42	.013	1.06	21.99
	SR \approx DR	-1.41	4.46	.755	-0.12	-2.11
	SU \approx DU	-2.15	5.23	.685	-0.16	-4.15
	DR > DU	13.87	4.23	.003	1.20	20.43
Anthro.	SR > SU	0.40	0.10	<.001	0.51	14.04
	SR \approx DR	-0.09	0.08	.262	-0.11	-3.23
	SU \approx DU	0.04	0.10	.685	0.05	1.59
	DR > DU	0.54	0.09	<.001	0.64	18.05
Animacy	SR > SU	0.60	0.10	<.001	0.84	18.07
	SR \approx DR	0.00	0.07	1.000	0.00	0.00
	SU \approx DU	-0.13	0.09	.178	-0.19	-4.63
	DR > DU	0.47	0.10	<.001	0.64	14.27
Likeability	SR > SU	0.63	0.10	<.001	0.93	16.81
	SR \approx DR	0.00	0.08	.965	-0.01	-0.10
	SU \approx DU	-0.03	0.09	.772	-0.04	-0.81
	DR > DU	0.60	0.10	<.001	0.86	16.22
Perceived Intelligence	SR > SU	1.19	0.13	<.001	1.77	27.63
	SR \approx DR	0.01	0.07	.881	0.02	0.25
	SU \approx DU	-0.02	0.12	.831	-0.03	-0.80
	DR > DU	1.15	0.13	<.001	1.53	26.87
Perceived Safety	SR > SU	0.45	0.09	<.001	0.78	12.67
	SR \approx DR	0.05	0.09	.565	0.09	1.52
	SU \approx DU	-0.05	0.08	.528	-0.09	-1.74
	DR > DU	0.34	0.11	.004	0.53	9.78

supporting H5. In the unreliable conditions, participants rated the ASM's future reliability lower than in the reliable conditions, which partially supported H6. In the conditions where the ASM made an error, PR scores were lower than scores in the nonerror conditions. Participant confidence in their assessment of the PR was also significantly lower in the error conditions than in the nonerror conditions. After witnessing an ASM error, participants rated their confidence in their assessment lower than when they had not witnessed an ASM error, partially supporting H8.

1) *Effect of Errors on Participant Perception of Reliability and Confidence:* The effect of ASM reliability on participant evaluations of PR and confidence in those evaluations was also examined for potential carry-over effects using independent

TABLE III
DESCRIPTIVE STATISTICS AND BETWEEN-CONDITION COMPARISON RESULTS FOR RELIABILITY PROJECTIONS, AND CONFIDENCE RATINGS, SHOWN BY EVENT TYPE

Event	Cond	M	SD	Between-Condition Comparison		
				p	d_s	% Diff.
PR	Begin	NE	3.77	.872	-0.02	-0.32
		E	3.78			
	Error	NE	3.88	<.001	2.57	50.04
		E	1.94			
	1st Follow	NE	3.80	.001	0.46	7.89
		E	3.50			
	2nd Follow	NE	3.93	.011	0.35	3.32
		E	3.80			
Conf in PR	Begin	NE	4.48	.807	0.03	0.52
		E	4.46			
	Error	NE	4.63	.001	0.44	6.37
		E	4.33			
	1st Follow	NE	4.53	.001	0.44	7.79
		E	4.17			
	2nd Follow	NE	4.73	.002	0.43	5.78
		E	4.46			

$N = 112$ in No Error (NE) condition, $N = 109$ in Error (E) condition.

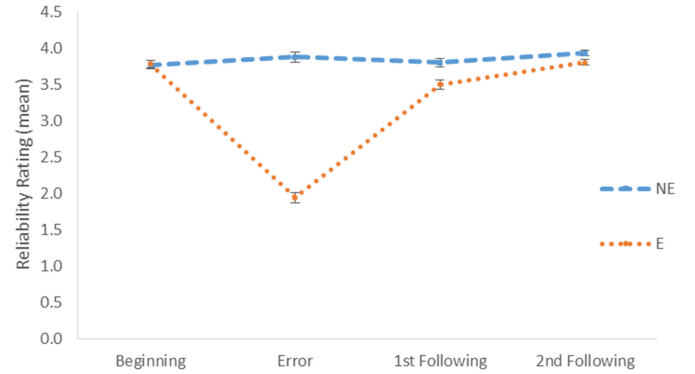


Fig. 5. Participant PR rating by event, sorted by whether the participant witnessed an ASM error during that event.

samples t-tests. There were four scenarios; each scenario was comprised of six events.

The event scores were grouped as such: beginning (all events from beginning of scenario up to the first error event), error (all events where the ASM made an error), 1st Following (the first nonerror event following an error event), and 2nd Following (the second nonerror event following an error event). Then the scores were averaged across scenarios, and the differences in scores were evaluated by whether the participant witnessed an error during that event (E; SU and DU conditions) or not (NE; SR and DR conditions). Witnessing an ASM error influenced participants' PR and confidence throughout the rest of the scenario (see Table III).

The beginning event in each scenario was error free, and there was no difference in assessments between the NE (no error witnessed) and E (error witnessed) groups (see Fig. 5). During the error event, there was a significant difference between the two groups, as was expected. Although the events following the

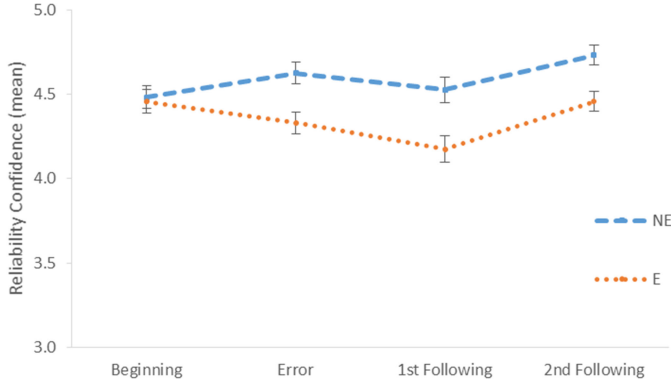


Fig. 6. Participant confidence in reliability ratings by event, sorted by whether the participant witnessed an ASM error during that event.

error event were error-free, the PR scores from those who had witnessed an error continued to be significantly lower than those who had not.

There was no difference in participant confidence between the NE (no error witnessed) and E (error witnessed) groups for the beginning event (see Fig. 6). During the error event, the NE group reported significantly higher confidence ratings than the E group. This difference in confidence ratings was consistent through the 1st and 2nd Following events. Participants who previously witnessed an error (E) reported reduced confidence in their evaluations of the ASM's projected accuracy and reliability than those that did not witness an error (NE).

Confidence scores for participants who had witnessed an error were then examined to determine whether ASM information level had any effect. Repeated-measures ANOVAs revealed no significant difference in confidence scores between participants who were in the surface-level information condition compared to those in the in-depth information condition for confidence in reliability ratings (Wilks' $\Lambda = 0.952$, $F(4, 104) = 1.30$, $p = 0.275$, and $\eta^2 = 0.048$).

D. Functional Trust Survey

Overall participant trust in the ASM was assessed via the functional trust survey. There was concern that participant fatigue affected outcomes in later scenarios, as the survey responses after the first trial were nuanced throughout the length of the survey, while the survey responses for subsequent trials were increasingly straight lined. Therefore, to mitigate the influence of potential respondent fatigue, only trust data for the first scenario were assessed [42] across participants.

There was a significant difference in overall participant trust between the experimental conditions. Planned comparisons indicated this difference was due to agent reliability, not varying information levels (see Table II), thus supporting H10 but not H9. When surface-level information was offered, overall trust was 18% higher in the reliable condition than the unreliable condition. When more in-depth information was provided, overall trust was 15.5% higher in the reliable condition than in the unreliable condition.

Participant trust in the ASM was also assessed along the four functions of automation using one-way ANOVAs with planned comparisons. There was a significant difference in trust in the ASM's ability to perform each of the automation functions between the experimental conditions (see Table II). Although the assessment was not statistically significant for Function B (integrating and displaying information), the effect sizes for the overall test and the planned comparisons were large, indicating there is most likely a difference in participant trust on this factor between conditions. Information level had no significant effect on operator trust, except in Function A (collecting and/or filtering information). Although the test did not show statistical significance, the effect sizes indicate that in-depth information does bolster trust, particularly when the ASM is unreliable. ASM reliability had a significant effect on participant trust across all automation functions. Participant trust in the ASM was consistently lower when the ASM was unreliable.

E. Humanness

Information level did not influence participant perceptions of the ASM, contrary to H11 (see Table II). Agent reliability did make a significant difference in participant perceptions of the ASM, supporting H12. When the ASM was reliable, participants anthropomorphized the agent more than when it was unreliable. Participants also rated the ASM as more animate, likable, intelligent, and safer to work with when the agent was reliable than when it was not.

IV. DISCUSSION

The purpose of this paper was to examine the interactive effects of agent transparency and agent reliability. To date, most research on agent transparency has focused on its utility in perfectly reliable systems. However, it is known that unreliable automation severely impacts operator performance and perceptions of a system. This paper examined whether increasing an agent's transparency would mitigate the undesirable effects of unreliable automation.

The primary purpose of the target detection task was to act as a secondary task that would maintain participant engagement with the ASM. This task was not intended to be challenging; however, some difference in performance due to varying information level or reliability was still expected. The findings suggest that access to in-depth agent SAT information or reduced agent reliability does not distract participants enough to influence their concurrent task performance (in this case, target detection efficiency). Considered together with the other findings of this study, this suggests that when the human teammate's tasks are independent of the agent, the agent's reliability and transparency could have limited (if any) influence on the human's task performance, while still having a significant impact on the human's perceptions of the agent.

Neither agent reliability nor information level affected the participant perceived workload. Prior studies have found that perceived workload did not increase because of increased transparency when working with reliable agents [6]; these findings demonstrate this may also be true when the agent is unreliable.

This outcome may be beneficial for designers and developers seeking to support more transparent interaction between agents and humans.

Participants' SA1 (perception) and SA2 (comprehension) scores were uniformly high throughout the scenarios, as were their self-confidence ratings of SA1 and SA2 responses. These high scores could be a demonstration of the effectiveness of the SAT framework in relaying information to the participant that supported their SA1 and SA2 of the agent. Alternatively, it could also indicate that the in-depth information did not contribute further to supporting SA beyond the surface level.

Participants' PR scores differed, depending on whether they had witnessed an ASM error. In the unreliable conditions, participants rated the ASM's future reliability lower than in reliable conditions. Prior research has shown that users may begin their work with an unfamiliar system with overly high expectations of the systems' performance (positivity bias; [45], [46]), and witnessing an error causes them to overcorrect their expectations, resulting in lower assessments of reliability than warranted [47]. It was expected that in-depth information from the ASM would mitigate these effects by allowing participants to understand precisely where the ASM "reasoning" failed. However, that was not the case, as the information level appeared to have no influence on participants' projections of ASM reliability.

There were carry-over effects from witnessing the ASM error. Once an error occurred, participants continued to rate the ASM as less reliable even when the ASM did not make any additional errors. These lowered ratings did improve as the ASM continued to demonstrate reliable behavior, and appeared to be trending to a point where they would once again be similar to those in the reliable conditions. These results are similar to findings that users will correct their low-reliability assessments of a system over time, so long as the user witnesses no further errors [47], indicating that the process for improving PR may be similar to trust repair [34].

Participants who witnessed an ASM error rated their confidence in their assessment lower than those who had not witnessed an ASM error. At the beginning of the scenarios, when no ASM error had occurred, there was no difference in reported confidence between participants in the two reliability conditions. Once the ASM made an error, those in the unreliable conditions reported significantly lower confidence in their predictions of ASM reliability than their reliable-condition counterparts. Surprisingly, this difference in confidence was consistent throughout the remainder of the scenario, i.e., even when no further errors occurred, participants who had witnessed an ASM error continued to report low confidence in their ability to assess and predict agent reliability.

Prior research has shown that participants report higher confidence in their responses when agreeing with a decision aid, rather than when they disagree [48], because consistently agreeing with the automation allows participants to assess the automation's reliability apart from their own. When the ASM initially erred, the participants reported lower reliability. However, on subsequent events when the ASM did not err, and they continued to rate its reliability as lowered, participants were most likely aware of their error in rating the reduced reliability. Confidence

in one's ability to correctly assimilate information, forming accurate comprehension and projection is fundamental to SA [26]. Even though the participants' assessment of the ASM's future reliability increased as the error event became more distant, their continued awareness of their inability to predict the ASM's reliability appeared to undermine their confidence in those assessments. While subjective confidence is not a reliable indicator of the validity of one's judgments [49], persons who lack confidence tend to be overly cautious, hesitant in decision-making, and slow to enact action [24]. As such, a reduction in one's confidence in decision-making could lead to undesirable results in a military setting.

Consistent with the findings of previous studies on automation reliability and trust [9], [21], [50], reliability, not information level, influenced participant trust. Participants also reported higher trust in the ASM when it was reliable for each of the four automation functions [41] than when it was unreliable. Information level did not affect participant trust in the ASM for the automation functions, except for the "collecting and/or filtering information" function, where participants with in-depth information reported higher trust than those with only surface-level information, in the (respectively) reliable and unreliable conditions. This finding suggests that in-depth knowledge of the underlying reasons for an agent's actions and reasoning may bolster trust in the agent and mitigate the effects of unreliable automation, at least so far as specific functions are concerned [51], [52]–[54].

Participant perceptions of the ASM were affected by agent reliability, but not information level. When the ASM displayed unreliable behavior, participants rated it as having less human-like qualities (e.g., less animate, likable, intelligent, and safer to work with) than the reliable ASM. Prior research has shown that participants rate an autonomous agent as more animate, likable, intelligent, and safer when it provides information supporting all three SAT levels [5], but the addition of in-depth information did not change participants' attitudes. We can infer that more information supporting transparency does not necessarily change the aforementioned attitude, so agent interface designers must use a deft hand when determining how much information the agent provides its human teammate, especially when they want to influence the relationship that humans have with that agent.

A. Limitations

This paper had several limitations. First, the taskload of the environmental monitoring tasks (i.e., target detection and event identification) was low, which most likely contributed to the lack of differences in task performance, workload, and SA. Second, participants were asked to take the functional trust survey after each scenario, but only the first survey was analyzed. While this single interaction provided a useful snapshot of participants' trust in the agent, this approach cannot reveal the change in trust over time that a multiple interaction approach could.

Finally, agent transparency (i.e., information level) had little-to-no influence on the study findings. This outcome was unexpected, and several potential factors may have contributed to this outcome. It is possible that the additional information was

not prominent enough, making it difficult for the participants to realize when it changed. This potential issue could be addressed in future studies by incorporating multimodal cues that would draw attention to the updated information [57]. It is also possible that the additional information provided to the participants was not needed, as the ASM's responses to events were clearly understood, rendering the additional information superfluous. In this study, the ASM's responses and underlying reasons were limited, so it is possible that the participants were able to identify the error and deduce the underlying reason based on their observations of the soldier team and environment. Finally, the methodology of this study may not have been sensitive enough to tease out the effects of transparency from a more powerful manipulation of system reliability [22]. However, the current results should not be interpreted as suggesting that transparency information is not useful, as the utility of transparency information is context dependent. Indeed, a recent real-world observational study [56] revealed that understandability of the automated system had the greatest impact on operators' trust in the system.

B. Future Research

Further investigation into the interaction between agent reliability and agent transparency is needed. Future research should explore how agent transparency interacts with agent reliability in a higher workload setting. Better understanding as to which tasks require a more in-depth understanding of the agent's reasoning, and how to discern what that depth would entail, is also needed. Perhaps more importantly, future research should explore ways to deliver transparency information, based on the tasking requirements. The delivery mechanisms could include other modalities besides visual. Multimodal communications and bidirectional transparency based on the SAT framework could be fruitful research areas [7].

V. CONCLUSION

Humans working with autonomous robots in simple, low workload environments may not have the same SAT needs as those in environments that are more dynamic. This study demonstrated that even when SAT information does not directly support task performance, it influences the human perceptions of the robot. Participants trusted in the agent's abilities to collect and filter information when given evidence of that activity, which suggests that future attempts to facilitate transparent human-agent interaction may benefit from reminding participants of the work the agent is doing "under the hood." Otherwise, visualizing explanatory elements of the agent's decision-making process, in situations where those decisions are not directly related to the human's actions, may not benefit human performance.

Agent reliability may be a stronger influence on the human's perceptions of the agent than agent transparency, although both are important for human-robot interaction [55]. Agent errors had a profound and lasting effect on the human teammates' perception of the agent and their confidence in their assessments. While their undesirable assessments of the agent's reliability appeared to diminish as the agent continued to display reliable behavior, their reduced confidence in their assessment of the

agent's reliability did not. Methods to restore the human teammates' confidence in their assessments of the robot should be explored, as it is crucial to appropriate continued use of the robot teammate.

ACKNOWLEDGMENT

The authors wish to thank Dr. D. Barber, M. Barnes, A. Watson, J. English, and J. Hart for their input and contributions to this project. The views and conclusions contained in this paper are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory, or the U.S. Government.

REFERENCES

- [1] "Autonomy research pilot initiative," Department of Defense, 2013. [Online]. Available: <http://www.acq.osd.mil/chieftechologist/arpi.html>
- [2] J. B. Lyons, "Being transparent about transparency: A model for human-robot interaction," presented at the 2013 AAAI Spring Symp. Ser., Palo Alto, CA, USA, 2013.
- [3] J. Y. C. Chen, K. Procci, M. W. Boyce, J. L. Wright, A. Garcia, and M. J. Barnes, "Situation awareness-based agent transparency," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, USA, Rep. ARL-TR-6905, 2014.
- [4] M. W. Boyce, J. Y. C. Chen, A. R. Selkowitz, and S. G. Lakhmani, "Agent transparency for an autonomous squad member," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, USA, Rep. ARL-TR-7298, 2015.
- [5] A. R. Selkowitz, S. G. Lakhmani, C. N. Larios, and J. Y. C. Chen, "Agent transparency and the autonomous squad member," *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 60, no. 1, pp. 1319–1323, 2016.
- [6] J. E. Mercado, M. A. Rupp, J. Y. C. Chen, M. J. Barnes, D. Barber, and K. Procci, "Intelligent agent transparency in human-agent teaming for multi-UxV management," *Human Factors*, vol. 58, no. 3, pp. 401–415, 2016, doi: [10.1177/0018720815621206](https://doi.org/10.1177/0018720815621206).
- [7] J. Y. C. Chen, S. G. Lakhmani, K. Stowers, A. R. Selkowitz, J. L. Wright, and M. J. Barnes, "Situation awareness-based agent transparency and human-autonomy teaming effectiveness," *Theor. Issues Ergonom. Sci.*, vol. 19, no. 3, pp. 259–282, 2018, doi: [10.1080/1463922X.2017.1315750](https://doi.org/10.1080/1463922X.2017.1315750).
- [8] K. Stubbs, D. Wettergreen, and P. H. Hinds, "Autonomy and common ground in human-robot interaction: A field study," *IEEE Intell. Syst.*, vol. 22, no. 2, pp. 42–50, Mar./Apr. 2007, doi: [10.1109/MIS.2007.21](https://doi.org/10.1109/MIS.2007.21).
- [9] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004, doi: [10.1518/hfes.46.1.50.30392](https://doi.org/10.1518/hfes.46.1.50.30392).
- [10] J. B. Lyons and P. R. Havig, "Transparency in a human-machine context: Approaches for fostering shared awareness/intent," in *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*. New York, NY, USA: Springer, 2014, pp. 181–190.
- [11] A. W. Evans, "Safe operations of unmanned systems for reconnaissance in complex environments-army technology objective (SOURCE ATO) field experimentation observations and soldier feedback," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, USA, Rep. ARL-TN-0488, 2012.
- [12] M. R. Endsley, "Autonomous horizons: System autonomy in the air force—A Path to the Future, volume I: Human-autonomy teaming," Dept. Air Force, Washington, D.C., USA, Rep. AF/ST TR 15-01, 2015.
- [13] J. L. Wright, J. Y. C. Chen, and M. J. Barnes, "Human-automation interaction for multiple robot control: The effect of varying automation assistance and individual differences on operator performance," *Ergonomics*, vol. 61, no. 8, pp. 1033–1045, 2018, doi: [10.1080/00140139.2018.1441449](https://doi.org/10.1080/00140139.2018.1441449).
- [14] S. G. Lakhmani, J. Y. C. Chen, J. L. Wright, A. R. Selkowitz, and M. R. Schwartz, "Agent transparency for an autonomous squad member: Uncertainty and projected outcomes," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, USA.
- [15] M. Klenk, M. Molineaux, and D. W. Aha, "Goal-driven autonomy for responding to unexpected events in strategy simulations," *Comput. Intell.*, vol. 29, pp. 187–206, 2013, doi: [10.1111/j.1467-8640.2012.00445.x](https://doi.org/10.1111/j.1467-8640.2012.00445.x).
- [16] A. R. Selkowitz, C. N. Larios, S. G. Lakhmani, and J. Y. C. Chen, "Displaying information to support transparency for autonomous platforms," in *Advances in Human Factors in Robots and Unmanned Systems*. New York, NY, USA: Springer, 2017, pp. 161–173.

- [17] M. P. Linegang *et al.*, "Human-automation collaboration in dynamic mission planning: A challenge requiring an ecological approach," *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 50, no. 23, pp. 2482–2486, 2006, doi: [10.1177/154193120605002304](https://doi.org/10.1177/154193120605002304).
- [18] S. E. McBride, W. A. Rogers, and A. D. Fisk, "Understanding human management of automation errors," *Theor. Issues Ergonom. Sci.*, vol. 15, no. 6, pp. 545–577, 2014, doi: [10.1080/1463922X.2013.817625](https://doi.org/10.1080/1463922X.2013.817625).
- [19] R. Parasuraman, R. Molloy, and I. L. Singh, "Performance consequences of automation-induced 'complacency'," *Int. J. Aviation Psychol.*, vol. 3, no. 1, pp. 1–23, 1993.
- [20] C. D. Wickens and S. R. Dixon, "The benefits of imperfect diagnostic automation: A synthesis of the literature," *Theor. Issues Ergonom. Sci.*, vol. 8, no. 3, pp. 201–212, 2007, doi: [10.1080/14639220500370105](https://doi.org/10.1080/14639220500370105).
- [21] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997.
- [22] J. L. Wright, J. Y. C. Chen, M. J. Barnes, and P. A. Hancock, "Agent reasoning transparency: The influence of information level on automation-induced complacency," U.S. Army Res. Lab., Aberdeen Proving Ground, MD, USA, Rep. ARL-TR-8044, 2017.
- [23] M. R. Endsley, "Final reflections: Situation awareness models and measures," *J. Cogn. Eng. Decis. Making*, vol. 9, no. 1, pp. 101–111, 2015, doi: [10.1177/1555343415573911](https://doi.org/10.1177/1555343415573911).
- [24] F. M. Lichacz, "An examination of situation awareness and confidence within a distributed information-sharing environment," *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 50, no. 3, pp. 344–348, 2006.
- [25] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995.
- [26] M. R. Endsley and W. M. Jones, "Situation awareness information dominance and information warfare," Endsley Consulting, Marietta, GA, USA, Rep. AL/CF-TR-1997-0156, 1997.
- [27] B. McGuinness, *Quantitative Analysis of Situational Awareness (QUASA): Applying Signal Detection Theory to True/False Probes and Self-Ratings*. Bristol, U.K.: Adv. Technol. Center, BAE Syst., 2004.
- [28] P. Kroft and C. D. Wickens, "Displaying multi-domain graphical database information: An evaluation of scanning, clutter, display size, and user activity," *Inf. Des. J.*, vol. 11, no. 1, pp. 44–52, 2002, doi: [10.1075/idx.11.1.06kro](https://doi.org/10.1075/idx.11.1.06kro).
- [29] M. C. Lohrenz, J. G. Trafton, M. R. Beck, and M. L. Gendron, "A model of clutter for complex, multivariate geospatial displays," *Human Factors*, vol. 51, no. 1, pp. 90–101, 2009, doi: [10.1177/0018720809333518](https://doi.org/10.1177/0018720809333518).
- [30] E. E. Jones and J. R. Kelly, "Why am I out of the loop?" Attribution influence responses to information exclusion," *Personality Social Psychol. Bull.*, vol. 36, no. 9, pp. 1186–1201, 2010, doi: [10.1177/0146167210380406](https://doi.org/10.1177/0146167210380406).
- [31] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, K, "Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust," *Proc. 10th Annu. ACM/IEEE Int. Conf. Human-Robot Interact.*, 2015, pp. 141–148.
- [32] M. Lewis, "Designing for human-agent interaction," *AI Mag.*, vol. 19, no. 2, pp. 67–78, 1998.
- [33] N. Haslam, "Dehumanization: An integrative review," *Personality Social Psychol. Rev.*, vol. 10, no. 3, pp. 252–264, 2006.
- [34] E. J. deVisser, R. Pak, and T. H. Shaw, "From 'automation' to 'autonomy': The importance of trust repair in human-machine interaction," *Ergonomics*, vol. 61, pp. 1–19, 2018, doi: [10.1080/00140139.2018.1457725](https://doi.org/10.1080/00140139.2018.1457725).
- [35] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, 1988, doi: [10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
- [36] D. G. Jones and D. B. Kaber, "Situation awareness measurement and the situation awareness global assessment technique," in *Handbook of Human Factors and Ergonomics Methods*, N. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, Eds. Boca Raton, FL, USA: CRC Press, 2004, pp. 421–428.
- [37] P. M. Salmon *et al.*, "Measuring situation awareness in complex systems: Comparison of measures study," *Int. J. Ind. Ergonom.*, vol. 39, no. 3, pp. 490–500, 2009, doi: [10.1016/j.ergon.2008.10.010](https://doi.org/10.1016/j.ergon.2008.10.010).
- [38] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ, USA: Prentice-Hall, 2003.
- [39] M. R. Endsley, "Situation awareness misconceptions and misunderstandings," *J. Cogn. Eng. Decis. Making*, vol. 9, no. 1, pp. 4–32, 2015, doi: [10.1177/1555343415572631](https://doi.org/10.1177/1555343415572631).
- [40] J. Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cogn. Ergonom.*, vol. 4, no. 1, pp. 53–71, 2000, doi: [10.1207/S15327566IJCE0401_04](https://doi.org/10.1207/S15327566IJCE0401_04).
- [41] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Syst., Man Cybern., Part A, Syst. Humans*, vol. 30, no. 3, pp. 286–297, May 2000.
- [42] V. N. O'Reilly-Shah, "Factors influencing healthcare provider respondent fatigue answering a globally administered in-app survey," *PeerJ*, vol. 5, 2017, Art. no. e3785.
- [43] K. Stowers, N. Kasdaglis, M. Rupp, J. Y. C. Chen, D. Barber, and M. J. Barnes, "Insights into human-agent teaming: Intelligent agent transparency and uncertainty," in *Advances in Human Factors in Robots and Unmanned Systems*. New York, NY, USA: Springer, 2017, pp. 149–160.
- [44] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *Int. J. Social Robot.*, vol. 1, no. 1, pp. 71–81, 2009, doi: [10.1007/s12369-008-0001-3](https://doi.org/10.1007/s12369-008-0001-3).
- [45] J. T. Cacioppo and G. G. Berntson, "Relationship between attitudes and evaluative space: A critical review with emphasis on the separability of positive and negative substrates," *Psychological Bull.*, vol. 115, no. 3, pp. 401–423, 1994, doi: [10.1037/0033-2909.115.3.401](https://doi.org/10.1037/0033-2909.115.3.401).
- [46] J. T. Cacioppo, W. L. Gardner, and G. G. Berntson, "Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space," *Personality Social Psychological Rev.*, vol. 1, no. 1, pp. 3–25, 1997, doi: [10.1207/s15327957pspr0101_2](https://doi.org/10.1207/s15327957pspr0101_2).
- [47] M. T. Dzindolet, L. G. Pierce, H. P. Beck, and L. A. Dawe, "Misuse and disuse of automated aids," *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 43, no. 3, pp. 339–343, 1999.
- [48] D. A. Wiegmann, "Agreeing with automated diagnostic aids: A study of users' concurrence strategies," *Human Factors*, vol. 44, no. 1, pp. 44–50, 2002, doi: [10.1518/0018720024494847](https://doi.org/10.1518/0018720024494847).
- [49] D. Kahneman and G. Klein, "Conditions for intuitive expertise: A failure to disagree," *Amer. Psychologist*, vol. 64, no. 6, pp. 515–526, 2009, doi: [10.1037/a0016755](https://doi.org/10.1037/a0016755).
- [50] E. J. de Visser and R. Parasuraman, "Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload," *J. Cogn. Eng. Decis. Making*, vol. 5, no. 2, pp. 209–231, 2011, doi: [10.1177/1555343411410160](https://doi.org/10.1177/1555343411410160).
- [51] B. M. Muir, "Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems," *Ergonomics*, vol. 37, no. 11, pp. 1905–1922, 1994.
- [52] M. T. Dzindolet, S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck, "The role of trust in automation reliance," *Int. J. Human-Comput. Stud.*, vol. 58, no. 6, pp. 697–719, 2003, doi: [10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7).
- [53] J. E. Mercado, M. A. Rupp, J. Y. C. Chen, D. Barber, K. Procci, and M. J. Barnes, "Effects of agent transparency on multi-robot management effectiveness" U.S. Army Res. Lab., Aberdeen Proving Ground, MD, USA, Rep. ARL-TR-7466, 2015.
- [54] K. A. Hoff and M. Bashir, M, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human Factors*, vol. 57, no. 3, pp. 407–434, 2015, doi: [10.1177/0018720814547570](https://doi.org/10.1177/0018720814547570).
- [55] K. E. Schaefer, J. Y. C. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Human Factors*, vol. 58, no. 3, pp. 377–400, 2016, doi: [10.1177/0018720816634228](https://doi.org/10.1177/0018720816634228).
- [56] N. Balfe, S. Sharples, and J. R. Wilson, "Understanding is key: An analysis of factors pertaining to trust in a real-world automation system," *Human Factors*, vol. 60, no. 4, pp. 477–495, 2018, doi: [10.1177/0018720818761256](https://doi.org/10.1177/0018720818761256).
- [57] J. Y. C. Chen and P. I. Terrence, "Effects of tactile cueing on concurrent performance of military and robotics tasks in a simulated multitasking environment," *Ergonomics* vol. 51, no. 8, pp. 1137–1152, 2008.