



ARL-TR-9893 • MAR 2024



Human–Agent Teaming with Learning-Capable Agents

by Julia L Wright, Shan G Lakhmani, Giovanna Camacho,
Justin Lee, and Jaquelyn Shreck

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

NOTICES

Disclaimers

The research reported herein was sponsored by the US Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) and was accomplished under Cooperative Agreement W911NF-19-2-0251 with the University of Central Florida. The views and conclusions contained in this report are those of the authors and should not be interpreted as representing the official policies or positions, either expressed or implied, of ARL, the US Government, or the University of Central Florida. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.



Human–Agent Teaming with Learning-Capable Agents

Julia L Wright and Shan G Lakhmani

DEVCOM Army Research Laboratory

Giovanna Camacho

Oak Ridge Associated Universities

Justin Lee and Jaquelyn Shreck

University of Central Florida Institute for Simulation and Training

REPORT DOCUMENTATION PAGE

1. REPORT DATE		2. REPORT TYPE		3. DATES COVERED					
March 2024		Technical Report		<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%;">START DATE</td> <td style="width: 50%;">END DATE</td> </tr> <tr> <td>1 September 2019</td> <td>30 December 2021</td> </tr> </table>		START DATE	END DATE	1 September 2019	30 December 2021
START DATE	END DATE								
1 September 2019	30 December 2021								
4. TITLE AND SUBTITLE									
Human-Agent Teaming with Learning-Capable Agents									
5a. CONTRACT NUMBER		5b. GRANT NUMBER		5c. PROGRAM ELEMENT NUMBER					
W911NF-19-2-0251									
5d. PROJECT NUMBER		5e. TASK NUMBER		5f. WORK UNIT NUMBER					
ARL 18-200									
6. AUTHOR(S)									
Julia L Wright, Shan G Lakhmani, Giovanna Camacho, Justin Lee, and Jaquelyn Shreck									
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER					
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-FD Orlando, FL 32826				ARL-TR-9893					
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		11. SPONSOR/MONITOR'S REPORT NUMBER(S)				
12. DISTRIBUTION/AVAILABILITY STATEMENT									
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.									
13. SUPPLEMENTARY NOTES									
ORCID IDs: Julia L Wright, 0000-0003-3026-1538; Shan G Lakhmani, 0000-0001-6052-439X									
14. ABSTRACT									
In this human-agent teaming experiment, participants took the role of a robotic combat vehicle gunner working with a simulated, learning-capable target identification agent to classify potential threats encountered while patrolling through a simulated environment. Participants completed four trials. The agent would classify persons as encountered and the participant reviewed the agent classification and submitted the report. When the agent could not complete a classification, it inferred needed information from the participant's input to later update its programming. Transparency of agent learning was manipulated in three trials. In the fourth trial, the participant directly "taught" the agent rather than allow it to infer the needed information. Within-subjects repeated-measures analysis showed learning transparency supported performance, workload, trust, and perceptions of the agent. However, in the human-directed learning condition, task performance, workload, performance satisfaction, trust in the agent, and perception of the agent's reliability, animacy, likeability, and intelligence all worsened compared to the learning from inferred information conditions. Overall, these findings indicate that transparency of agent learning is an important design feature to include to support effective human-agent teaming, but human-directed learning should be fielded with care, as the deleterious results could jeopardize mission effectiveness.									
15. SUBJECT TERMS									
artificial intelligence/machine learning, AI/ML, intervention learning, autonomy, agent transparency, trust, uncertainty, workload, Godspeed, individual differences, Human-Autonomy Teaming, Humans in Complex Systems									
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT		18. NUMBER OF PAGES				
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU		149				
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED							
19a. NAME OF RESPONSIBLE PERSON				19b. PHONE NUMBER (Include area code)					
Julia L Wright				(407) 208-3348					

STANDARD FORM 298 (REV. 5/2020)

Prescribed by ANSI Std. Z39.18

Contents

List of Figures	vi
List of Tables	viii
Acknowledgments	xi
Executive Summary	xii
1. Introduction	1
2. Background	2
2.1 Autonomy/Agents	2
2.2 Human–Agent Teams	3
2.3 Situation Awareness and Transparency	4
2.4 Workload	5
2.5 Trust in Automated Systems	5
2.6 Human Perceptions of Agents	6
2.7 Individual Differences	6
2.7.1 Attentional Control	7
2.7.2 Intolerance of Uncertainty	7
2.7.3 Personality	7
2.8 Current Study	8
2.8.1 Research Objective	11
2.8.2 Hypotheses	12
3. Method	13
3.1 Participants	13
3.2 Apparatus (Sim and Hardware)	13
3.3 Questionnaires and Tests	15
3.3.1 Demographics Questionnaire	15
3.3.2 Ishihara Color Vision Test	15
3.3.3 Workload	15

3.3.4	Trust in Automation	16
3.3.5	Anthropomorphic Tendencies	16
3.3.6	Perceived Attentional Control	16
3.3.7	Personality	16
3.3.8	Intolerance of Uncertainty	17
3.4	Experimental Design	17
3.4.1	Independent Variables	17
3.4.2	Dependent Measures	18
3.4.3	Procedure	20
4.	Results	20
4.1	Data Analysis	20
4.2	Task Performance	21
4.2.1	Correct Classifications	21
4.2.2	Response Times	22
4.3	Workload	23
4.3.1	Overall Scores	23
4.3.2	Mental Demand	24
4.3.3	Physical Demand	24
4.3.4	Temporal Demand	24
4.3.5	Effort	25
4.3.6	Frustration	25
4.3.7	Performance Satisfaction	25
4.4	Situation Awareness	25
4.5	Perceived Reliability	26
4.6	Trust	27
4.7	Godspeed	27
4.8	Individual Differences	29
4.8.1	Perceived Attentional Control	29
4.8.2	Uncertainty	32
4.8.3	Personality	36
5.	Discussion	42
5.1	Limitations	46
5.2	Future Directions	46

6. Conclusion	46
7. References	48
Appendix A. Threat Classification Factors and Training	57
Appendix B. Demographics Questionnaire	62
Appendix C. Ishihara Color Test	64
Appendix D. NASA-Task Load Index (TLX)	66
Appendix E. Trust in Automation	68
Appendix F. Godspeed Questionnaire Survey	70
Appendix G. Perceived Attentional Control Survey	72
Appendix H. Mini-International Personality Item Pool (IPIP)	74
Appendix I. Intolerance of Uncertainty	76
Appendix J. Situation Awareness Global Assessment Technique (SAGAT) Probes	78
Appendix K. University of Central Florida (UCF) COVID Information Sheet and Screening Questionnaire	80
Appendix L. Data Tables	84
List of Symbols, Abbreviations, and Acronyms	132
Distribution List	134

List of Figures

Fig. 1	Target classification task interface (right-hand monitor). Windows are (clockwise from right) indirect camera vision feeds (red border), agent reasoning and transparency (yellow border), snapshot of selected person (green border), and list of identified persons (blue border). For additional details see Section 3.4.	9
Fig. 2	Environment awareness interface (left-hand monitor). Windows are (clockwise from right) top-down map of the area showing vehicle and its route (red border), route selection logic (blue border), and command chat window (yellow border). For additional details see Section 3.4.....	11
Fig. 3	Participant station with two monitors, keyboard, and mouse	14
Fig. 4	Example left monitor showing map of the area, planned route, command communication module, and route comparator.....	14
Fig. 5	Example right monitor showing camera feed, threat classification, and potential threat list.....	15
Fig. 6	Correct rejections by trial.....	21
Fig. 7	Comparison of correct IDs within trial by task load (high vs. low) ...	22
Fig. 8	Overall response times (average seconds) by transparency condition. Bars denote standard error (SE).....	22
Fig. 9	NASA-TLX scores by factor, over transparency condition. Bar denotes SE. MD = mental demand, PD = physical demand, TD = temporal demand.....	24
Fig. 10	Perceived reliability scores by transparency condition. Higher scores indicate higher perceived reliability. Bars denote SE.	26
Fig. 11	Jian TIA scale scores across trials	27
Fig. 12	Godspeed questionnaire scores by condition.....	28
Fig. 13	High PAC vs. low PAC response times for correct and incorrect identifications.....	30
Fig. 14	NASA-TLX scores by low vs. high PAC	30
Fig. 15	Jian TIA scale scores by low vs. high PAC	31
Fig. 16	Task performance (correct IDs) by prospective anxiety and inhibitory anxiety	33
Fig. 17	Response times by high/low prospective and inhibitory anxiety.....	33
Fig. 18	Selected NASA-TLX scores by high/low prospective anxiety. MD = mental demand, EFF = effort, FRUST = frustration, PERF SAT = performance satisfaction. Bars denote SE.	34

Fig. 19	Selected NASA-TLX scores by high/low inhibitory anxiety. MD - mental demand, EFF = effort, FRUST = frustration, PERF SAT = performance satisfaction. Bars denote SE.	34
Fig. 20	Selected Godspeed questionnaire scores by high/low prospective anxiety. Bars denote SE.	35
Fig. 21	Selected Godspeed questionnaire scores by high/low inhibitory anxiety. Bars denote SE.	36
Fig. 22	Range of scores for each personality factor. The measure's potential range of scores is shown as the left-most entry. Each personality factor's scores are shown as overall, then how they were sorted into high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.	37
Fig. C-1	Example of Ishihara color test image.....	65

List of Tables

Table 1	Reasoning by learning transparency condition matrix.....	17
Table 2	Descriptive statistics for personality factor data.....	36
Table 3	Task performance high/low group outcomes at a glance. X indicates high group had better performance; O indicates high group had worse performance. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.	38
Table 4	NASA-TLX high/low group outcomes at a glance. X indicates high group had lower scores; O indicates high group had higher scores. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.	39
Table 5	Sagat-Probe high/low group outcomes at a glance. X indicates high group had higher scores; O indicates high group had lower scores. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.	40
Table 6	Jian TIA scale high/low group outcomes at a glance. X indicates high group had higher scores; O indicates high group had lower scores. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.	41
Table 7	Godspeed questionnaire high/low group survey outcomes at a glance. X indicates high group had higher scores; O indicates high group had lower scores. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.	42
Table L-1	Descriptive statistics for performance measures.....	85
Table L-2	Paired t-test results for performance measures	85
Table L-3	Descriptive statistics by task load for each transparency condition ...	85
Table L-4	T-test results high vs. low task load outcomes within each task load condition	86
Table L-5	Descriptive statistics for NASA-TLX results	86
Table L-6	Paired T-test results for NASA-TLX scores	86
Table L-7	Situation awareness (SA) and perceived reliability scores descriptive statistics. LTL – low task load. HTL – high task load.	86
Table L-8	Paired t-test results for situation awareness and perceived reliability results	87
Table L-9	Paired t-test results for SA and perceived reliability scores by task load (low LTL vs. high HTL)	87

Table L-10	Descriptive statistics for trust and godspeed questionnaires.....	87
Table L-11	Paired t-test results for Trust and Godspeed questionnaires between transparency conditions	87
Table L-12	Performance statistics by high/low perceived attentional control ...	88
Table L-13	Workload factor scores by high/low perceived attentional control .	89
Table L-14	Situation awareness scores by high/low perceived attentional control	91
Table L-15	Jian TIA scale scores by high/low perceived attentional control	92
Table L-16	Godspeed survey scores by high/low perceived attentional control	93
Table L-17	Performance measures by high/low prospective anxiety	94
Table L-18	Performance measures by high/low inhibitory anxiety	95
Table L-19	Workload factor scores by high/low prospective anxiety	96
Table L-20	Workload factor scores by high/low inhibitory anxiety	98
Table L-21	Situation awareness scores by high/low prospective anxiety	100
Table L-22	Situation awareness scores by high/low inhibitory anxiety	101
Table L-23	Jian TIA scale scores by high/low prospective anxiety	102
Table L-24	Jian TIA scale scores by high/low inhibitory anxiety.....	102
Table L-25	Godspeed survey scores by high/low prospective anxiety	103
Table L-26	Godspeed survey scores by high/low inhibitory anxiety	104
Table L-27	Performance statistics by high/low extroversion scores	105
Table L-28	Performance statistics by high/low agreeableness scores.....	106
Table L-29	Performance statistics by high/low conscientiousness scores	107
Table L-30	Performance statistics by high/low neuroticism scores	108
Table L-31	Performance statistics by high/low openness scores	109
Table L-32	Workload (NASA-TLX) statistics by high/low extroversion scores.....	110
Table L-33	Workload (NASA-TLX) statistics by high/low agreeableness scores.....	112
Table L-34	Workload (NASA-TLX) statistics by high/low conscientiousness scores.....	114
Table L-35	Workload (NASA-TLX) statistics by high/low neuroticism scores.....	116
Table L-36	Workload (NASA-TLX) statistics by high/low openness scores ..	118
Table L-37	Situation awareness scores by high/low extroversion	120
Table L-38	Situation awareness scores by high/low agreeableness	121
Table L-39	Situation awareness scores by high/low conscientiousness.....	122

Table L-40	Situation awareness scores by high/low neuroticism.....	123
Table L-41	Situation awareness scores by high/low openness.....	124
Table L-42	Jian trust survey scores by high/low extroversion	125
Table L-43	Jian trust survey scores by high/low agreeableness.....	125
Table L-44	Jian trust survey scores by high/low conscientiousness	125
Table L-45	Jian trust survey scores by high/low neuroticism	126
Table L-46	Jian trust survey scores by high/low openness	126
Table L-47	Godspeed survey scores by high/low extroversion.....	127
Table L-48	Godspeed survey scores by high/low agreeableness	128
Table L-49	Godspeed survey scores by high/low conscientiousness	129
Table L-50	Godspeed survey scores by high/low neuroticism.....	130
Table L-51	Godspeed survey scores by high/low openness	131

Acknowledgments

The authors wish to thank Jonathan Coole for his contributions to this project.

Executive Summary

This experiment investigated how transparency of agent reasoning and learning within a human-agent team influenced the human team member's task performance, situation awareness (SA), workload, trust in the agent, and perceptions of the agent. Participants worked with a learning-capable aided target recognition agent to identify potential threats as the robotic combat vehicle traversed a simulated urban environment. The simulated agents' reasoning transparency (what factors were used in determining threat) and learning transparency (how participant-supplied information was interpreted by the agent) were manipulated between trials, and task load (i.e., low vs. high) was varied within each trial. There were two levels of reasoning transparency: Opaque (no agent reasoning information) and transparent (agent reasoning information available). Learning transparency had three levels: implicit learning (the agent gave no information as to how it used participant-supplied information to update its algorithm), explicit learning (the agent shared how it interpreted supplied information for its algorithm update), and human-directed learning (the participant specifically told the agent how to interpret the supplied information). Within-subjects analyses examined differences in task performance, cognitive workload, SA, trust, and attitudes due to differing transparency conditions. Individual difference factors (i.e., attentional control, personality, intolerance to uncertainty) were also examined as between-subject variables.

Participants completed four trials, one in each transparency condition. Results indicate that increasing transparency of the agent's learning process supports task performance, lowers response times, and lowers workload. There was little effect on participant trust in the agent, their SA1 and SA2 scores, and attitudes pertaining to the agent. However, in the human-directed learning condition, task performance, workload, performance satisfaction, trust in the agent, and perception of agent reliability, animacy, likeability, and intelligence all worsened compared to the conditions where the agent learning was from inferred information. There were also differences due to individual differences factors. High attentional control individuals had better task performance, SA, trust, and perceptions of the agent than the low attentional control individuals; however, the low attentional control group reported better performance satisfaction in the human-directed learning condition. Prospective anxiety (PA) and inhibitory anxiety (IA) were examined to better explain the effects of uncertainty on performance. There was little distinction between the groups; both high PA and high IA individuals had fewer correct individual differences but faster response times, lower trust, lower mental demand, and perceived the agent as less likeable, intelligent, and safe than their low PA and low IA counterparts. In the human-directed learning condition the high PA group

had greater frustration and better performance satisfaction than the low PA group, while the high IA group had lower performance satisfaction than the low IA group. There were varied differences due to personality factors, most notably, high conscientiousness individuals appeared to bond better with the agent, as evidenced by their increased trust and high perceptions of agent animacy, intelligence, and likeability.

Overall, these findings indicate that transparency of agent learning, wherein the agent shares how it is interpreting user-supplied inputs to update its algorithms, is an important design feature to include to support effective human-agent teaming. However, human-directed learning, wherein the human must teach the agent during the mission, should be fielded with care, as the deleterious results could jeopardize mission effectiveness.

1. Introduction

Vehicles of the future can be autonomous, partially autonomous, fully manual, or have combinations of these states. Part of successfully fielding these vehicles is recognizing where the human operator and the autonomous systems each fall short and what actions can be taken to enhance their effectiveness. Vehicle autonomy allows for the human operators to attend to multiple tasks, or in the case of military vehicles, place fewer warfighters in harm's way. However, when humans are working with autonomous teammates on joint tasking, it becomes difficult to keep the human in the loop of control, thus aware of the agent's actions. When the operator is unaware of the reasoning behind the intelligent agent's suggestions, complacent behavior may occur, resulting in misuse of the agent or distrust of the agent can lead to lack of use even when appropriate (Wright et al. 2017). Increasing the transparency of the intelligent agent has been shown to be effective at keeping the human operator in the loop of control, thus improving task performance, situation awareness (SA), and reducing complacent behaviors, while not increasing their perceived workload (Kilgore and Voshell 2014; Mercado et al. 2015; Wright et al. 2017). However, it is still unknown how well agent transparency will support the human's performance when the agent is capable of "learning" by updating its programming autonomously.

Advances in artificial intelligence (AI) and machine learning (ML) continue to increase not only the abilities of these agents to conduct tasks autonomously, but also to conduct tasks of ever-increasing complexity and difficulty. These gains have given rise to learning-capable agents (LCAs), agents that will adapt their behavior over time as their models of the world are updated by events encountered. LCAs can adapt to new information, updating their goals and reasoning without the knowledge of their human counterpart. Thus, when human-agent teams (HATs) include LCAs, within-team communications become crucial to shared awareness and team effectiveness. While the effects of agent transparency on operator performance have been studied with agents that are incapable of independently updating their goals and/or reasoning, little is known about outcomes related to human-agent teaming with an LCA. The goal of this study is to examine how these advances in AI/ML affect the human teammate, specifically, how transparency of agent reasoning and agent learning affects team performance.

2. Background

2.1 Autonomy/Agents

Autonomous systems (i.e., intelligent agents) can be either embodied (e.g., robots) or embedded (i.e., software) and, unlike automation, are able to exercise some amount of choice regarding their actions using information they have gathered from their environment (Russell and Norvig 2009). To be considered an independent agent, the autonomy first must have agency, the ability to make decisions, select courses of action, and direct its activities (Lyons et al. 2018; Wynne and Lyons 2018; O'Neill et al. 2020). These intelligent agents can also be rational, interacting with their environment to achieve a specific goal and then use specific criteria to measure their success (Wooldridge and Jennings 1995; David and Nielsen 2016).

LCAs use their experiences to update the preprogrammed models that direct their actions and goals. These agents can be taught via a variety of methods (e.g., a priori training, real-time learning, on-the-job training). Training can be completely guided by human interaction (e.g., learning by demonstration) or can be through exploration, wherein the human offers reinforcement or advice, but the training does not require the human's undivided attention (Breazeal and Thomaz 2008). In real-time or on-the-job learning, the agent learns using similar strategies as a human, where learning is a part of all activities. In those cases, the human operator serves as the teacher that will guide the agent providing feedback, directing them to specific tasks/actions, and manipulating the surrounding environment (Breazeal and Thomaz 2008). Regardless of the learning method, these agents can only learn within a predefined scope, only perform tasks for which it has received prior training, and typically require vast amounts of data to update their algorithms (e.g., learn something new).

One example of an LCA, image recognition agents, use camera images to understand their surroundings and inform their actions. These agents are trained for real-world tasks via presentation of a series of images prior to deployment (off-line learning). During training, images are labeled and categorized, either a priori, real time by the agent, or post hoc with human assistance. However, it is not possible to generate labeled images of all potential threats and situations a priori, hence, the need for the agent to be able to continue learning after deployment (Kim et al. 2007).

There are several potential solutions to the issue of continual training for image recognition agents. For example, learning could occur after deployment by having the agent periodically review the targets that the agent missed but the human identified, essentially repeated off-line learning. This method would allow the agent

to update its models based on all new information gathered since the previous learning session; however, the targets must not only be identified but classified as well. This repeats the iterative a priori training process step-by-step—the agent classifies what it can based on the new information, then the human reviews and classifies what the agent could not.

Alternatively, on-the-job training, wherein the human teaches the agent, could be used to improve an agent’s learning and performance while the team is conducting tasks in the field. There is utility to the agent being able to learn real time after deployment (i.e., lifelong learning). For example, having the human disambiguate unclear stimuli by classifying “partially identified” images (e.g., the agent recognizes a human, but cannot classify it as threat or no threat) in real time has the potential to create a more agile, responsive battle force in addition to reducing the time required for additional off-line learning. In this situation, a man holding a cylindrical object may be holding either a rifle or a broom. These two situations require vastly different responses from the HAT, so correct identification is critical to mission success. In this study, the simulated agent will use information supplied by the human during missions to update its knowledge base off-line (i.e., learn).

2.2 Human-Agent Teams

Human-agent teaming (HAT) introduces agents to complete concurrent or parallel subtask objectives of the overall mission set. This concept matches the definition of a team where, “two or more members work interdependently toward a common goal” (O’Neill et al. 2022). There are many benefits of autonomous agent collaboration, but it is imperative to plan for the hurdles Soldiers must face in acclimating to this new teammate. Beyond the ability to collaborate, autonomous agents would need to also have a degree of agency, where the agents would be proactive in completing independent actions of their own (Lyons et al. 2018). The criteria for HAT would not be met with autonomous agents lacking a sense of agency as humans would be less likely to view the agent as a teammate (Lyons et al. 2018) and more like a tool. Teaming with an autonomous agent does not always result in improved task performance and SA (Chen et al. 2010; Mercado et al. 2016). At high levels of autonomy or when task load is high, the human may lose sight of what the agent is doing or why, resulting in an out-of-the-loop situation, encouraging automation bias, complacent behavior, or other undesirable behaviors (Parasuraman and Riley 1997; Wright et al. 2017; Lin et al. 2020).

2.3 Situation Awareness and Transparency

SA is defined as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (Endsley 1988). This three-level model operationalizes SA into three levels: perceiving, understanding, and predicting (Endsley 1995). Maintaining adequate SA can become tremendously taxing in an up-tempo environment, even more so when the individual is teamed with an autonomous agent. When human operators have difficulty understanding the agent’s state and actions, they will question the agent’s accuracy and effectiveness (Bitan and Meyer 2007; Seppelt and Lee 2007; Stanton et al. 2007). The human operator can become overwhelmed by the amount of information, leaving them out of the loop and resulting in complacent behavior (Chen et al. 2010; Wright et al. 2017, 2021). Even when the agent notifies the human operator of their actions, the human has difficulty maintaining awareness of actions they were not involved in (Parasuraman et al. 2000). Clearly, maintaining sufficient SA, of both the task environment and the autonomous agent, is crucial for the human operator to make appropriate decisions and calibrate their trust in the system (Lee and See 2004; de Visser et al. 2014).

Studies have shown that when the human operator understands the agent actions and intention, they remain in the loop of control, improving performance and reducing deleterious outcomes such as automation-induced complacency (Parasuraman et al. 2000; Lyons 2013; Schaefer et al. 2017; Wright et al. 2017; Lin et al. 2020). Collaboration between the team members is essential to developing systems that will complement the strengths and weaknesses of one another to leverage capabilities in task execution (Ososky et al. 2014). Bidirectional communications support collaborative efforts by creating and maintaining transparency within the team (Chen et al. 2014, 2018). Transparency is defined as “able to be seen through” or “easy to notice or understand” (Ososky et al. 2014). Agent transparency is defined as “the descriptive quality of an interface pertaining to its abilities to afford an operator’s comprehension about an intelligent agent’s intent, performance, future plans, and reasoning process” (Chen et al. 2014). Thus, for bidirectional team transparency to occur, each team member shares the responsibility for maintaining transparency by sharing information (as it pertains to themselves and their tasking) (Lyons 2017; Chen et al. 2018). Team members benefit from bidirectional transparency, as it supports the development of a shared mental model within the team and facilitates decision making (Chen et al. 2018; Mathieu et al. 2000). This shared mental model is essential for organizing team knowledge and team performance (Mathieu et al. 2000).

2.4 Workload

Workload, according to attention resource theory, is the relationship between the amount of resources required by a task and the amount of resources the human has available (Yeh and Wickens 1988; Parasuraman et al. 2008). As such, it is an individualistic construct, as “two people performing the same task (or two systems used by the same person) can generate identical behavioral and performance outputs, yet one person may have plenty of attention left to allocate to concurrent tasks, whereas the other does not” (Parasuraman et al. 2008). Assessments of an individual’s cognitive workload can be performance-based, subjective, or physiological. The goal in assessing workload is to ensure that the benefits to the individual outweigh the costs incurred from the working environment.

Overall, workload with regards to HAT depends heavily on the levels of automaticity and transparency. Studies have shown that in a target detection-type task a human operator can supervise between 4 and 9+ agents, depending on the environmental demands and the agent’s level of autonomy (Wang et al. 2009). With each individual agent completing their own task, the human operator’s workload increases to levels where they are unable to maintain command and control. Increasing the autonomy of the agents can allow them to be unattended for longer durations, which would allow an operator to control more agents (Wang et al. 2009). In conjunction with increasing agent automaticity, summarized data can help to make communication within operations more efficient (Parasuraman et al. 2000). Information that supports transparency can help increase SA and trust without impacting workload (Selkowitz 2016).

2.5 Trust in Automated Systems

Trust is a multidimensional construct with cognitive and emotional components (Riley 1989) that lead to specific behaviors (Taylor 1990). Trust has been defined as “a psychological state comprising the intention to accept vulnerability based upon the positive expectations of the intentions or behavior of another” (Rousseau et al. 1998) and “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” (Lee and See 2004). The concepts of trust hold true across human-to-human trust, human-machine trust, and trust in general (Jian et al. 2000); however, trust is not a fixed characteristic. Rather, it is an impermanent willingness to be subjected to the outcomes of reciprocity and betrayal (Evans and Revelle 2008). In thinking of trust from this mindset, it is understandable that when trust is violated it is not regained easily (Deutsch 1958; Muir 1988).

In HATs, calibrated trust is key to maintaining effective teaming (Parasuraman and Riley 1997). This is difficult to achieve, as operator trust in the agent is influenced by both the operator’s implicit and explicit attitudes, experience, and propensity to trust, and the agent’s reliability, design transparency, and performance, sometimes across different contexts, conditions, or tasks (Merritt et al. 2013; Chen et al. 2014). Poor calibration is often the result of an individual misinterpreting the capabilities of the agent. Transparency of the agents’ actions, intents, processes, and constraints support appropriate trust calibration (Lee and See 2004; Chen et al. 2014, 2018). The human is more trusting of an agent that has understandable algorithms and operations that govern behavior in favor of the operator’s end goals (Lee and See 2004). In this study, the agent is an LCA, as such it is able to update its goals and understanding of its environment based upon its experiences. As such, one query in this investigation is to determine if trust in this HAT is similar to that in other, static agent teams.

2.6 Human Perceptions of Agents

How the human perceives their robotic teammate is of interest in this study. Prior research has shown that, while users may perceive the autonomy as having some human-like traits (Nass and Moon 2000), they do make distinctions between humans and autonomy in trust (de Visser et al. 2016; Sanders et al. 2019), performance expectations (Merritt et al. 2015), and assigning intent to the autonomy (Lee and See 2004). There is evidence that when comparing the behavior of two entities, humans can discern between humans and agents (Wykowska et al. 2015).

However, there is evidence these biases are tempered by increased perception of humanness in the agent (de Visser et al. 2016). When the robot is perceived as more trustworthy, reliable, responsible, and more transparent with their reasoning and projected outcomes, the robot is also evaluated as more human-like (Hinds et al. 2004; Merritt et al. 2015; Selkowitz et al. 2016; Wright et al. 2020a). These interactions have been examined with autonomy with static programming. In this study we examined if these findings also pertain to LCAs.

2.7 Individual Differences

Multiple individual difference (ID) factors have been shown to influence human-agent teaming outcomes (Chen and Barnes 2012), particularly operator performance (Wright et al. 2018) and attitudes toward the agent (Wright et al. 2020b). Of particular interest to human-agent teaming paradigms are how well one

can focus their attention, handle uncertainty, and personality traits that may support (or hinder) effective teaming. As such the following ID factors were examined.

2.7.1 Attentional Control

High perceived attentional control (PAC) benefits persons in tasks that require attention shifting and focus (Derryberry and Reed 2002). In human–agent teaming paradigms that require multitasking and scanning the environment, high attentional control individuals outperformed their low attentional control counterparts (Chen and Barnes 2012; Wright et al. 2013). High PAC individuals report greater trust in the agent and confidence in their assessments of the agent than low PAC individuals, particularly when there is limited information available (Wright et al. 2020b).

2.7.2 Intolerance of Uncertainty

The tendency of an individual to consider that an unacceptable negative event may occur, regardless of the likelihood of such occurring, is referred to as intolerance of uncertainty (IUS; Carleton et al. 2007). This intolerance is a key factor in worry, state anxiety, and so on, and can influence current and future actions of the individual. High IUS individuals do not deal well with ambiguous situations, are more likely to deem unclear circumstances and information as threats and are likely to have their anxiety interfere with their problem-solving skills, resulting in inaction or avoidance behaviors (Dugas et al. 1997; Greco and Roger 2001, 2003; Nomura et al. 2007; Birrell et al. 2011).

2.7.3 Personality

Individual differences in personality may contribute in multiple and complex ways to the interactions between a human and an agent on a social level (Matthews et al. 2021). When the human has difficulty understanding the agent, they have difficulty maintaining an accurate mental model of the world, their team, and even their actions and goals. This inherent difficulty is expected to increase as technological advances are made, further exacerbating the problem.

While there are multiple models of personality, the “Big 5” dimensions of personality—extraversion, agreeableness, conscientiousness, neuroticism, and openness (McCrae and Costa 1997)—appear to have the greatest consensus among researchers. Extraversion encompasses assertiveness, excitability, sociability, and the degree in which people engage with the external world and has been positively correlated with the propensity to trust machines (Merritt and Ilgen 2008). High extraversion has been shown to support multitasking (Matthews 2008), impressions of agent intelligence (Hamacher et al. 2016), and feelings of closeness and bonding

with the agent (Syrdal et al. 2006; Salem et al. 2015;). Agreeableness as a personality trait describes persons who demonstrate prosocial behaviors, kindness, affection, altruism, and cooperative. Agreeableness has been found to positively correlate with higher trust in machines (Chien et al. 2016). Conscientiousness as a personality trait is expressed as ambitious, goal-oriented, organized, being direct and unflinching, and highly motivated to perform tasks well. It has shown to be a predictor of low trust in robots (Lin et al. 2020), and high conscientiousness operators may experience higher workload even in periods of lower demand. Neuroticism is associated with negative effects, which may include emotional instability, anxiety, anger, depression, and low self-esteem. High neuroticism individuals do not handle stress well, often report higher workload, and have poorer performance (Boyes and French 2010; Szalma and Taylor 2011). These individuals are sensitive to criticism and uncertainty (Servaas et al. 2013), both of which would negatively impact human-agent teaming outcomes. The openness personality trait describes a person’s curiosity of the world. Persons high in openness typically have many interests, are eager to learn, highly imaginative, and insightful. It is believed that high openness individuals would handle uncertainty well (Goff and Ackerman 1992) but may become frustrated if the agent’s reasoning is not clear (Weld and Bansal 2018).

2.8 Current Study

In the current experiment, participants will take the role of a robotic combat vehicle (RCV) gunner whose RCV is patrolling a simulated urban area identifying potential threats (i.e., target identification task). The RCV gunner shares the target identification task with a scripted “autonomous” agent whose classifications are predetermined (unbeknownst to the participant). The agent simulates a computer-vision-enabled aided target recognition (AiTR) system that uses feature-based characteristics of persons encountered in the environment along the RCV’s route to classify that person as either a potential threat or a non-threat. The agent indicates recognition and classification of persons by highlighting them in the indirect vision camera feed (Fig. 1). The participant reviews the agent’s classification and submits the report, which clears the buttons in preparation for the next classification. The participant can override the agent’s classification if desired. At times, the agent will recognize a human or object, but will be unable to classify it as a potential threat or not. When this occurs, the agent will indicate the incomplete classification and the gunner will complete the classification by indicating yes/threat or no/non-threat.



Fig. 1 Target classification task interface (right-hand monitor). Windows are (clockwise from right) indirect camera vision feeds (red border), agent reasoning and transparency (yellow border), snapshot of selected person (green border), and list of identified persons (blue border). For additional details see Section 3.4.

Targets are classified as either potential threat or not a threat using binary logic across five factors: human/not human, civilian/military, known affiliation apparent/no affiliation, in a group/not in a group, and armed/unarmed. The factors are shown on the interface by five icon pairs, and the agent indicates its understanding by highlighting which icon pair was used in its classification. When the agent cannot determine which of the factor’s icon pair is true, that pair remains dimmed. There are targets in the simulation that are designed as to appear to confuse the agent to lend support to the notion that the agent sometimes cannot complete a classification. For example, a person may be wearing camo pants but no other military accruements; the idea is that the agent cannot determine whether that person is a member of the military. The agent does not make errors in classification, only occasional incomplete classifications. See Appendix A for details on threat classification.

The agent will either share its reasoning for the classification in the robot dialogue window on the interface (transparent) or it will not (opaque). When agent reasoning is opaque, the icons indicating the reasoning factors are not present (see Fig. 1, agent reasoning and transparency window). Agent reasoning must be transparent (icons present) for learning transparency to be present. When agent learning is opaque, the agent will not reveal how the participant’s input will contribute to agent learning (implicit learning). When agent learning is transparent, once the participant indicates threat/non-threat, the agent backfills the missing reasoning

information, showing the participant how their input was interpreted to be used for agent learning (explicit learning). In the human-assisted learning condition, the participant fills in the missing reasoning information by clicking on the appropriate icon and then indicates threat or non-threat.

The participants' primary task will be working with the simulated agent, identifying potential threats in the environment. The threat identification task will consist of participants monitoring the vehicles' camera feeds, reviewing agent classifications of potential threats, and classifying threats when they are present and not previously classified by the agent. The utility of agent transparency will be examined as it pertains to the threat identification task and agent learning. In the current study, we will also manipulate participant task load. All participants will experience periods of high versus low task load, which will be manipulated by increasing the number of potential threats to be classified during specified periods.

In addition to the primary task, participants will be required to monitor the RCV's progress through the environment using the top-down map, and to maintain awareness of communications from their higher-level commander (see Fig. 2). The participant will monitor the simulated route-planning agent's recommendations for revised routes as the RCV proceeds along its patrol route. Events (e.g., threats, environmental hazards/obstacles, instructions from command) will occur that may require changing its route to avoid potentially hazardous situations or explore additional areas. The agent will suggest route revisions, displaying the competing plans side-by-side, and select the route based upon mission requirements. Icons indicating the potential threats and their classifications will also appear on the map. The participant will not be required participate in the route selection or communications tasks; however, they will be required to maintain SA of the route, the environment the RGV traverses, and all communications. The agent's route-planning task's transparency and difficulty will be unchanged throughout all scenarios.



Fig. 2 Environment awareness interface (left-hand monitor). Windows are (clockwise from right) top-down map of the area showing vehicle and its route (red border), route selection logic (blue border), and command chat window (yellow border). For additional details see Section 3.4.

2.8.1 Research Objective

This within-subjects study will assess the effects of transparency of agent learning on a target identification task. The transparency of how the participant's contribution to agent learning will be examined to see how it affects human performance, trust, and workload in a multitasking, dynamic environment. In the baseline condition, opaque reasoning and implicit learning, the agent will conduct the task without sharing its reasoning or querying reasoning for partially identified targets. In the transparent reasoning/implicit learning condition, the agent will share its reasoning but will not indicate how the participant's input is used in updating its underlying reasoning. In the transparent reasoning/explicit learning condition, the agent will share its reasoning and indicate how the participant's input is used in updating its underlying reasoning (backfill reasoning information based upon participant input). Finally, in the human-assisted learning condition, the agent will share its reasoning and the participant will complete the reasoning information by completing the missing information (clicking the appropriate icon) before identifying whether the target is a potential threat. The number of targets correctly identified and response time will be used to assess participant performance on the threat identification task. Task load will be manipulated via the threat identification task by varying the number of potential threats the agent needs assistance in identifying during specified periods of the scenario. Participants will also be required to maintain SA and will receive SA queries throughout each trial.

The current study investigates the effect of agent transparency of reasoning and learning on participants' performance, workload, SA, and perceptions of an autonomous agent. The effect of individual differences (i.e., PAC, uncertainty tolerance, and personality) on the participant's task performance will also be investigated. Based on the review of the literature, the outcomes expected are listed in Section 2.8.2.

2.8.2 Hypotheses

The hypotheses for this study were as follows:

- 1) When agent reasoning is transparent, compared to the opaque reasoning condition, there will be
 - 1.1 Increased correct classifications on the threat identification task
 - 1.2 Decreased time required to classify threats.
 - 1.3 Improved SA scores
 - 1.4 Reduced participant perceived workload
 - 1.5 Increased participant trust in the agent
 - 1.6 Higher evaluations of agent intelligence, animacy, and likeability.
- 2) When agent learning is explicit, compared to the implicit learning condition, there will be
 - 2.1 Increased correct classifications on the threat identification task
 - 2.2 Decreased time required to classify threats
 - 2.3 Improved SA scores
 - 2.4 Reduced participant perceived workload
 - 2.5 Increased participant trust in the agent
 - 2.6 Higher evaluations of agent intelligence, animacy, and likeability.
- 3) When the human assists in agent learning, compared to the explicit learning condition, there will be
 - 3.1 Increased correct classifications on the threat identification task
 - 3.2 Increased time required to classify threats
 - 3.3 Reduced SA scores
 - 3.4 Increased participant perceived workload
 - 3.5 Increased participant trust in the agent
 - 3.6 Higher evaluations of agent intelligence, animacy, and likeability.
- 4) When task load is high, compared to low task load, transparent agent reasoning will

- 4.1 Reduce performance on the threat identification task, as measured by number of threats identified
- 4.2 Increase complacent behavior (as indicated by reduced time to agree with agent classifications)
- 4.3 Reduce SA scores
- 5) When task load is high, compared to the low task load, explicit agent learning will
 - 5.1 Reduce performance on the threat identification task, as measured by number of threats identified
 - 5.2 Increase complacent behavior (as indicated by reduced time to agree with agent classifications)
 - 5.3 Reduce SA scores
- 6) There will be differential results on all dependent measures (i.e., threat identification performance, trust, workload, SA, and perceptions of the agent) due to individual difference factors (i.e., PAC, personality factors, response to uncertainty).

3. Method

3.1 Participants

A total of 25 participants between the ages of 18 and 40 were recruited from the general population. Five participants were excluded from the final analysis (i.e., one due to experimenter error and four due to simulator malfunction). As such, data from 20 participants were included in the final analysis, $Min_{age} = 18$ years, $Max_{age} = 28$ years, $M_{age} = 21.7$ years. Participants were compensated \$15 per hour. All participants were required to have acceptable color vision and were screened using the Ishihara Color Vision test (Ishihara 1917).

3.2 Apparatus (Sim and Hardware)

The study simulation was an urban environment developed in the Unreal 4 engine. A custom user interface software was presented to participants using a desktop computer with two 15-inch monitors (1680×1050 resolution), standard wireless keyboard, and three-button mouse (Fig. 3). The study was conducted in an indoor, climate-controlled laboratory/office space.



Fig. 3 Participant station with two monitors, keyboard, and mouse

The left monitor showed the route comparator, an overhead view map of the simulation environment, and a text display for “commander messages” (Fig. 4). The right monitor shows the simulated environment from the perspective of the RCV and buttons relevant to the identification and classification tasks (Fig. 5).

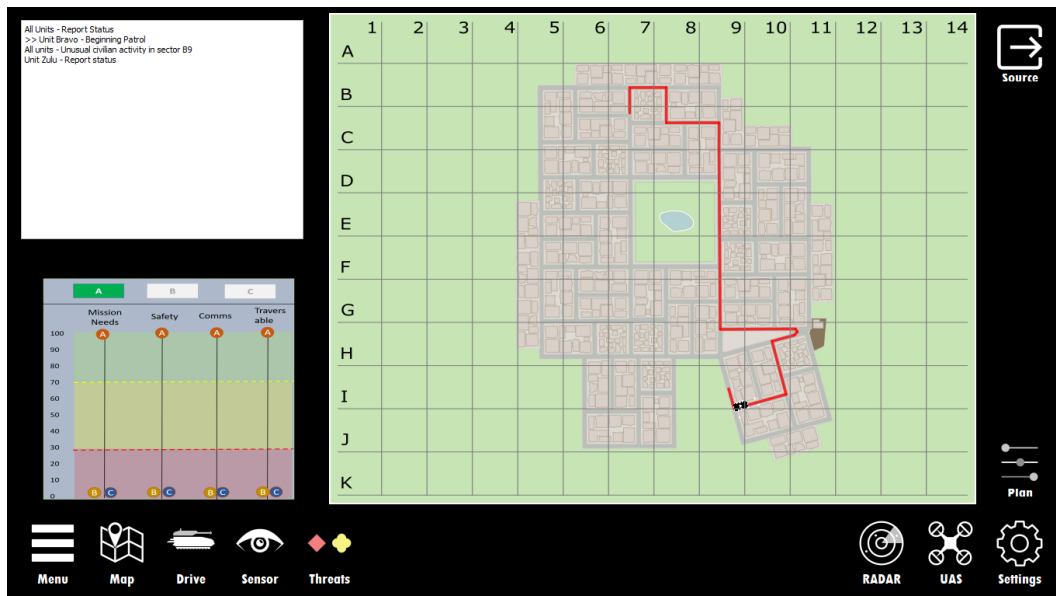


Fig. 4 Example left monitor showing map of the area, planned route, command communication module, and route comparator



Fig. 5 Example right monitor showing camera feed, threat classification, and potential threat list

3.3 Questionnaires and Tests

3.3.1 Demographics Questionnaire

A demographics questionnaire was administered before the training session began (see Appendix B). Information on the participants' age, gender, military experience, education, computer familiarity, and gaming experience was collected.

3.3.2 Ishihara Color Vision Test

The participants' color vision was assessed using the Ishihara Color Vision Test (1917), nine-test plate version (Appendix C). Since the displays used employ color-coded information, normal color vision was required to effectively interact with the system. Slides were delivered using a PowerPoint presentation. Participants were required to score 8/9 correctly to continue with the study.

3.3.3 Workload

Participants' subjective workload was assessed using the computerized version of the NASA-Task Load Index (TLX) questionnaire (Hart and Staveland 1988). The NASA-TLX is a self-report questionnaire of perceived demands in six areas: mental demand, physical demand, temporal demand, effort (mental and physical), frustration, and performance. Participants evaluated their perceived workload on these factors Likert-type scales ranging from 0 (low, good) to 100 (high, poor).

Overall workload was assessed using an average of all raw factor scores (Appendix D).

3.3.4 Trust in Automation

At the end of each scenario participants assessed their trust in the AiTR agent using a modified version of the TIA Scale (Jian et al. 2000). This scale comprises 12 items, such as “The AiTR is deceptive,” that the participant assesses using 7-point Likert scales (1 = not at all, 7 = extremely), wherein higher scores indicate more trust in the agent (Appendix E).

3.3.5 Anthropomorphic Tendencies

The Godspeed Questionnaire Series (GQS) consists of 24 questions that assess an individual’s perceptions of a robot on such attributes as anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety via a series of bipolar Likert scale evaluations (Bartneck et al. 2008). The GQS items (Appendix F) are rated on 5-point Likert scales with endpoints that are semantic differentials (e.g., awful/nice). Internal reliability consistency ranges from 0.702 (animacy) to 0.929 (anthropomorphism). Participants completed a Godspeed measure after each experimental trial.

3.3.6 Perceived Attentional Control

The PAC questionnaire (Derryberry and Reed 2002) was used to determine participants’ PAC by evaluating their perception of their attention focus and shifting. The Attentional Control Survey consists of 20 items scored on a 7-point Likert scale, with half of the items reverse-scored (Appendix G). The potential score range is 20–80 points, with higher scores indicating better attentional control. The scale has been shown to have good internal reliability ($\alpha = 0.88$).

3.3.7 Personality

The mini-International Personality Item Pool (IPIP) questionnaire (Donnellan et al. 2006) was used to determine an individual’s personality across five distinct factors. The mini-IPIP is a 20-item questionnaire that assesses five personality traits: extraversion, agreeableness, conscientiousness, neuroticism, and openness (Goldberg 1999; Donnellan et al. 2006). Participants evaluate each item using 5-point Likert scales, 1 = Very Inaccurate (doesn’t sound like me) to 5 = Very Accurate (sounds very much like me). Scores for each factor range from 4 to 20 (Appendix H). The internal consistencies of this survey were all above ($\alpha = 0.60$) and are well correlated with the long form of the IPIP.

3.3.8 Intolerance of Uncertainty

The Intolerance of Uncertainty Scale – Short Format (Appendix I; Buhr and Dugas 2002; Carelton et al. 2007) assesses the tendency of an individual to respond inappropriately to the possibility that a negative event may occur. The shortened survey is a 12-item questionnaire that assesses two factors, prospective anxiety and inhibitory anxiety, using 4-point Likert scales (Appendix I). Prospective anxiety is the fear and anxiety that is based on future events. Inhibitory anxiety is when actions or decisions are hindered by uncertainty. Each factor has high internal reliability, prospective anxiety (7 items, $\alpha = 0.85$) and inhibitory anxiety (5 items, $\alpha = 0.85$). The shortened version highly correlates with the original 27-item survey ($r = 0.96$).

3.4 Experimental Design

This study was a $2 \times 3 \times 2$ fractional-factorial design experiment. Within-subjects evaluations compared differences in performance and attributions regarding the agent across levels of transparency of agent reasoning and learning, and task load on a threat classification task. Between-subjects evaluations assessed how individual differences affected these outcomes.

3.4.1 Independent Variables

The independent variables were AiTR reasoning (reasoning transparency; opaque vs. transparent), AiTR learning (learning transparency; implicit, explicit, human-directed), and task load (task load: low, high). Since reasoning transparency and learning transparency are conveyed using the same icons, it is not possible to have transparent learning with opaque reasoning (see Table 1).

Table 1 Reasoning by learning transparency condition matrix

		Learning		
		Implicit	Explicit	Human-directed
Reasoning	Opaque	T1	...	...
	Transparent	T2	T3	T4

3.4.1.1 Transparency

In the baseline condition, Opaque Reasoning/Implicit Learning (T1), the agent conducts the task without sharing its reasoning for classifying persons. In the transparent reasoning/implicit learning (T2) condition, the agent shares its reasoning for classifying persons but does not indicate how the participant's input is used in updating its underlying reasoning. In the transparent reasoning/explicit learning (T3) condition, the agent shares its reasoning for classifying persons and

backfills the reasoning information based upon participant input to indicate what it inferred from the human's input. Finally, in the human-directed learning (T4) condition, the agent shares its reasoning for classifying persons and the participant completes any missing reasoning information before identifying whether the target is a potential threat. Participants completed four scenarios, one in each transparency combination condition. Scenarios were approximately 16 min long, and task load was varied in 4-min increments (e.g., high, low, high, low) by increasing/decreasing the event rate (all persons encountered).

3.4.1.2 Task Load

Task load was manipulated by the number of classifications the participant had to complete, 5 in each low task load (LTL) section, 10 in each high task load (HTL) section. The event rate stayed consistent at approximately 10 signals per minute. The AiTR could correctly classify most of the persons encountered; however, for some, its reasoning was incomplete, and the participant had to classify these events (unknowns) alone.

3.4.2 Dependent Measures

3.4.2.1 Threat Classification Task

During each scenario, participants worked with a simulated AiTR agent to classify persons in the environment as either potential threat or not a potential threat. Each scenario had roughly 175 events (persons) total, presentation rate of approximately 10/min. The AiTR would identify a person in the environment and make an initial classification. The participant had to review each classification and make their own determination, then submit the report. The participant could override the AiTR if they determined the AiTR classification was in error. Unbeknownst to the participant, the AiTR was perfectly reliable when classifying persons in the environment (i.e., if the AiTR completed a classification, it was correct). Performance on this task was assessed by number of correct classifications and response time.

3.4.2.2 Workload

After each mission, the NASA-TLX questionnaire was administered to assess the participants' perceived workload. Both global and individual factor workload scores were evaluated.

3.4.2.3 Situation Awareness

To assess the participant's current awareness of their environment and the AiTR, the Situation Awareness Global Assessment Technique (SAGAT) was used (Jones and Kaber 2004). SAGAT is a method where SA-related queries are administered to participants during predetermined freezes of the simulation during the task under analysis (Jones and Kaber 2004; Salmon et al. 2009; Endsley 2017). During each scenario, the simulation was paused four times. At each pause, the participant answered several SA1 and SA2 queries (Appendix J). Each scenario had 10 SA1 queries and 6 SA2 queries. Queries were scored as correct (1) or incorrect (0) and reported as mean scores, with higher scores indicating better SA.

3.4.2.4 Perceived Reliability

Each scenario also had four perceived reliability queries that assessed the participants' understanding of the AiTR agents' reliability. Unbeknownst to the participants, the agent made no mistakes (100% reliable). The classifications the agent could not complete had some correct information and then an unknown factor, hence the agent was still "reliable," it simply needed more information to complete the classification. While the participants were unaware of the agent's actual reliability, they were informed during training that missing information was not an error, it was simply part of improving the agent's understanding of persons in its environment (i.e., teaching the agent). It is important to understand how this need for further training influenced the participants' perception of the agent's reliability, as such, at each SAGAT probe the participants were asked to rate the agent's reliability. These were scored as Very reliable (no errors) = 2, Mostly reliable (few errors) = 1, Unreliable (many errors) = 0, and reported as a mean score (Range 0–2), with higher scores indicating higher perceived reliability.

3.4.2.5 Trust

After each scenario participants completed the modified Jian et al. (2000) TIA scale.

3.4.2.6 Anthropomorphic Tendencies

After each scenario, the Godspeed Questionnaire Series was administered to assess the participant's attributions of anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of the AiTR agent.

3.4.3 Procedure

On arrival at the lab, all potential participants were screened for COVID-19 using the University of Central Florida Covid Self-Checker (see Appendix K for information sheet and questions). Persons who did not pass the self-checker requirements were not allowed to enter the laboratory. Once the screening was complete, the participant was brought into the lab and briefed on the purpose of the study. They then signed the informed consent form and were screened for normal color vision using the Ishihara Color Vision Test (Ishihara 1917). After passing the color vision screening participants completed the demographics questionnaire, the PAC, the mini-IPIP questionnaire and the Intolerance of Uncertainty scale.

Participants received training via a self-paced PowerPoint slide deck. This training was self-guided, took around an hour to complete, and familiarized the participants with their task as well as the interface. The training included several quizzes in which the participant needed to achieve a score of 8/10 before they could continue with the rest of the training. If they scored lower than 80%, they were allowed to repeat the previous section and take the quiz again. They were allowed to ask questions if they needed clarification. After the training slides, participants completed a short practice scenario in which they could practice their task. After training, participants were allowed to have a short break before continuing with the experiment.

Each participant completed four scenarios, each in a different reasoning/learning transparency combination. During the scenarios, the simulation would pause several times and the SAGAT-style queries were presented. After each scenario, participants completed the NASA-TLX, Godspeed, and TIA surveys. After they completed four scenarios, the participant was debriefed and dismissed.

4. Results

4.1 Data Analysis

Data were analyzed using repeated-measures analysis of variances (ANOVAs) ($\alpha = 0.05$) and planned comparisons (T1 to T2, T2 to T3, T3 to T4). When appropriate, the Greenhouse-Geisser correction for sphericity violation is reported. Planned comparisons were conducted using paired t-tests, and Cohen's d effect sizes are reported for these analyses. Descriptive statistics and t-test result tables are shown in Appendix L.

4.2 Task Performance

4.2.1 Correct Classifications

Repeated-measures within-subjects ANOVA indicated there was a significant effect of agent transparency on overall correct classifications, Wilks' $\Lambda = 0.613$, $F(3, 17) = 3.57$, $p = 0.036$. Paired comparisons show that while the addition of agent reasoning transparency significantly increased the percentage of correct classifications ($M_{T1} = 91.3\%$, $SD = 4.8$; $M_{T2} = 93.8\%$, $SD = 3.85$; $t(19) = -2.44$, $p = 0.025$, $d_s = 0.55$) there was no difference due to agent learning transparency (see Appendix L Tables L-1 and L-2).

Participant classification errors comprised two types: incorrect classifications when the agent could not complete the classification, and incorrect rejections of the agent's classification. There were no significant differences due to transparency in the percent of correct classifications when the participant was completing an incomplete classification, Wilks' $\Lambda = 0.874$, $F(3, 17) = 0.82$, $p = 0.502$. There was a significant effect of reasoning transparency on incorrect rejections, Wilks' $\Lambda = 0.623$, $F(3, 17) = 3.42$, $p = 0.041$. While task performance overall was very good, paired comparisons showed there was a significant improvement in incorrect rejections (higher percent of correct classifications) due to the addition of reasoning transparency ($M_{T1} = 93.2\%$, $SD = 4.60$; $M_{T2} = 95.9\%$, $SD = 3.73$; $t(19) = -2.66$, $p = 0.016$, $d_s = 0.66$) (Fig. 6).

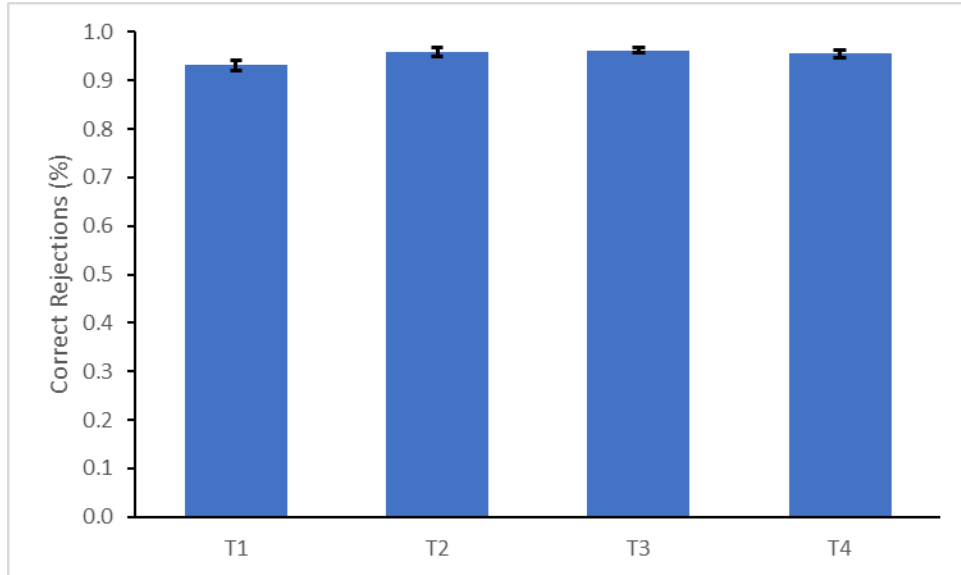


Fig. 6 Correct rejections by trial

There was a significant difference in overall correct classifications due to task load in all transparency conditions, T1 $d_s = 1.01$; T2 $d_s = 0.82$; T3 $d_s = 0.96$; T4 $d_s = 0.97$

(see Appendix L, Tables L-3 and L-4). Participants had fewer correct classifications (percent of total classifications) when task load was high compared to when task load was low (Fig. 7).

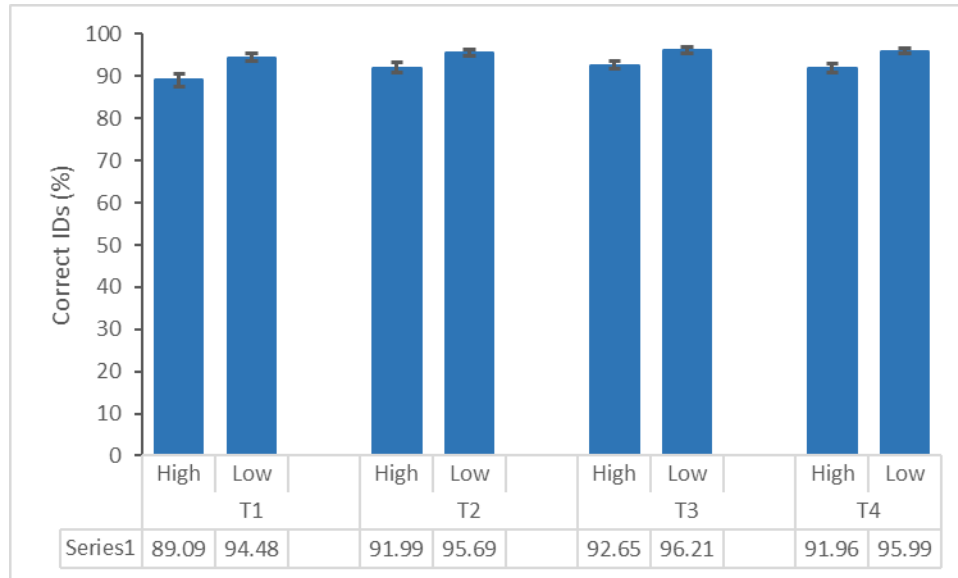


Fig. 1 Comparison of correct IDs within trial by task load (high vs. low)

4.2.2 Response Times

Participants had multiple targets onscreen at once and could review targets they had passed earlier in the trial for classification later. As such, response times were only examined in overall terms, as a per target analysis would not be informative. Repeated-measures within-subjects ANOVA indicated there was a significant effect of agent transparency on the overall response time, Wilks' $\Lambda = 0.163$, $F(1.52, 17) = 18.14$, $p < 0.001$ (Fig. 8).

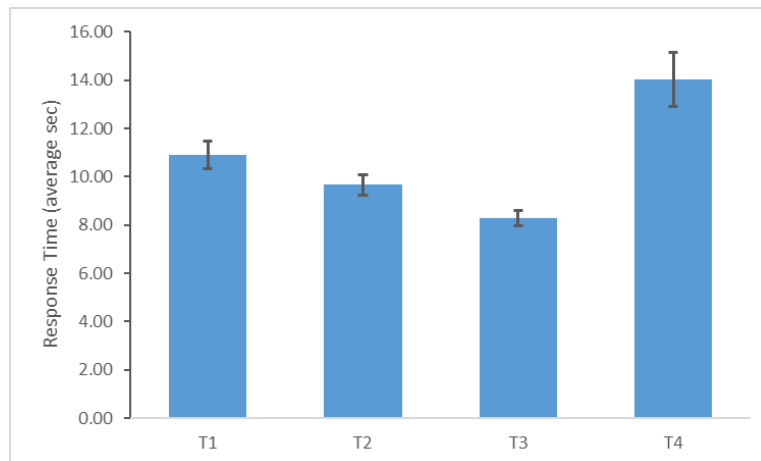


Fig. 8 Overall response times (average seconds) by transparency condition. Bars denote standard error (SE).

Paired comparisons show that the addition of agent reasoning transparency significantly reduced participant response time ($M_{T1} = 10.89$ s, $SD = 2.54$; $M_{T2} = 9.66$ s, $SD = 1.89$; $t(19) = 2.28$, $p = 0.034$, $d_s = 0.55$), and the addition of learning transparency further reduced response time ($M_{T2} > M_{T3}$; $M_{T3} = 8.29$ s, $SD = 1.39$; $t(19) = 4.36$, $p < 0.001$, $d_s = 0.83$). However, the change to human-directed learning significantly increased response time, ($M_{T3} < M_{T4}$; $M_{T4} = 14.03$ s, $SD = 4.95$, $t(19) = -5.95$, $p < 0.001$, $d_s = -1.58$). Response times were highest in the human-directed learning condition (T4).

There was no significant difference in response time due to task load in the reasoning or learning transparency conditions; however, in the human-directed learning condition (T4) response times when task load was high were significantly longer than response times when task load was low ($M_{LTL} < M_{HTL}$; $M_{LTL} = 12.08$ s, $SD = 3.33$; $M_{HTL} = 19.018$ s, $SD = 11.58$, $t(19) = 3.42$, $p = 0.003$, $d_s = 0.81$, see Appendix L, Tables L-3 and L-4).

4.3 Workload

See Appendix L, Tables L-5 and L-6, for workload descriptive statistics and paired t-test results, and Fig. 8 for comparison of survey results by condition.

4.3.1 Overall Scores

There was a significant effect of transparency condition on overall NASA-TLX scores, Wilks' $\Lambda = 0.440$, $F(2.09, 17) = 6.36$, $p = 0.004$ (Fig. 9). Paired t-tests revealed a marginally significant difference in overall workload due to reasoning transparency ($M_{T1} = 50.29$, $SD = 16.52$; $M_{T2} = 44.63$, $SD = 17.57$; $t(19) = 2.02$, $p = 0.058$, $d_s = 0.33$), showing the addition of reasoning information reduced overall cognitive workload. There was also a significant difference in overall workload due to human-directed learning ($M_{T3} = 41.33$, $SD = 16.69$; $M_{T4} = 53.46$, $SD = 16.91$; $t(19) = 4.54$, $p < 0.001$, $d_s = 0.72$), indicating that when the participant had to explicitly indicate what the missing information was there was a negative impact (increase) in overall cognitive workload. Learning transparency had no effect on overall workload.

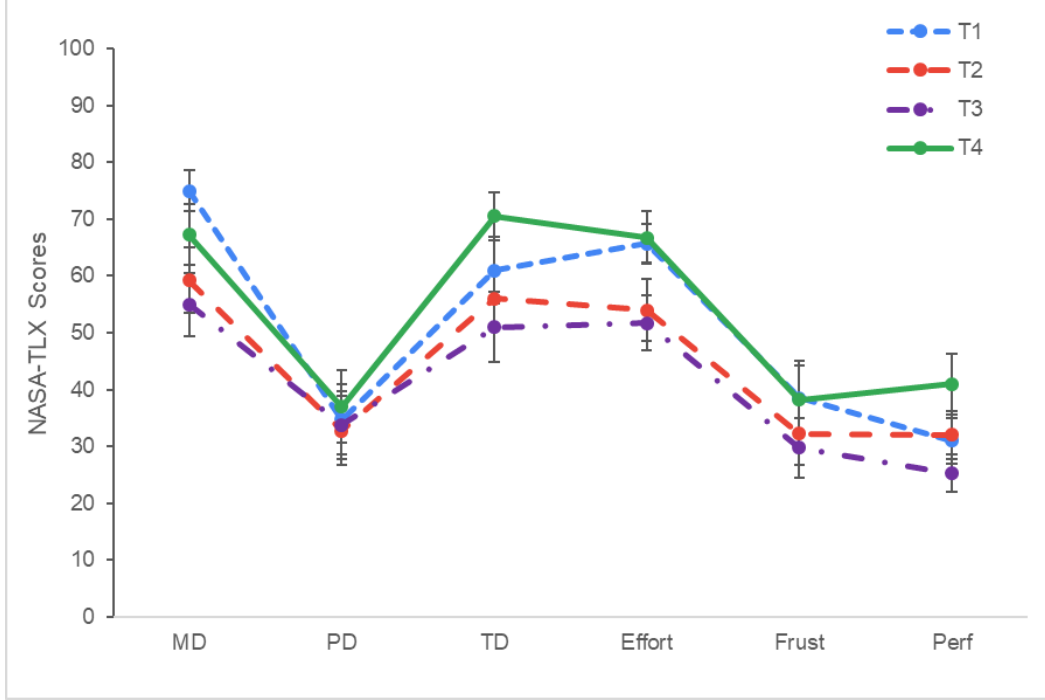


Fig. 9 NASA-TLX scores by factor, over transparency condition. Bar denotes SE. MD = mental demand, PD = physical demand, TD = temporal demand.

4.3.2 Mental Demand

There was a significant effect of transparency condition on NASA-TLX mental demand (MD) scores, Wilks' $\Lambda = 0.400$, $F(3, 17) = 8.49$, $p = 0.001$. Paired t-tests revealed MD significantly reduced when reasoning information was present ($M_{T1} = 75.00$, $SD = 16.38$; $M_{T2} = 59.25$, $SD = 25.56$; $t(19) = 4.55$, $p < 0.001$, $d_s = 0.73$) and increased when human-directed learning was needed ($M_{T3} = 55.00$, $SD = 24.71$; $M_{T4} = 67.25$, $SD = 24.03$; $t(19) = 3.39$, $p = 0.003$, $d_s = 0.50$). Learning transparency had no effect on MD scores. MD was highest in the opaque reasoning condition (T1).

4.3.3 Physical Demand

Repeated-measures ANOVA indicated there was no significant effect of transparency condition on NASA-TLX physical demand (PD) scores, Wilks' $\Lambda = 0.916$, $F(1.86, 17) = 0.43$, $p = 0.639$. Paired t-tests revealed no significant differences in PD due to reasoning or learning transparency, or human-directed learning.

4.3.4 Temporal Demand

There was a significant effect of transparency condition on NASA-TLX temporal demand (TD) scores, Wilks' $\Lambda = 0.524$, $F(2.21, 17) = 5.42$, $p = 0.007$. Paired t-tests

revealed no significant differences in TD due to reasoning or learning transparency; however, there was a significant increase in TD scores in the human-directed learning condition ($M_{T3} = 51.00$, $SD = 27.61$; $M_{T4} = 70.50$, $SD = 18.98$; $t(19) = 3.96$, $p = 0.001$, $d_s = 0.82$). TD was highest in the human-directed learning condition, T4.

4.3.5 Effort

There was a significant effect of transparency condition on NASA-TLX effort scores, Wilks' $\Lambda = 0.386$, $F(3, 17) = 9.03$, $p = 0.001$. Paired t-tests revealed reported effort significantly reduced when reasoning information was present ($M_{T1} = 65.75$, $SD = 15.16$; $M_{T2} = 54.00$, $SD = 24.31$; $t(19) = 2.66$, $p = 0.015$, $d_s = 0.58$) and increased when human-directed learning was needed ($M_{T3} = 51.75$, $SD = 21.42$; $M_{T4} = 66.75$, $SD = 20.98$; $t(19) = 5.43$, $p < 0.001$, $d_s = 0.71$). Learning transparency had no effect on effort scores. Effort was greatest in the human-directed learning condition, T4.

4.3.6 Frustration

Repeated-measures ANOVA indicated there was no significant effect of transparency condition on NASA-TLX frustration scores, Wilks' $\Lambda = 0.719$, $F(1.72, 17) = 1.50$, $p = 0.237$. Paired t-tests revealed no significant differences in frustration scores due to reasoning or learning transparency, or human-directed learning.

4.3.7 Performance Satisfaction

There was a significant effect of transparency condition on performance satisfaction scores, Wilks' $\Lambda = 0.400$, $F(3, 17) = 8.49$, $p = 0.001$. Paired t-tests revealed no significant difference in performance satisfaction scores due to reasoning transparency. Increasing learning transparency improved performance satisfaction ($M_{T2} = 32.00$, $SD = 18.74$; $M_{T3} = 25.25$, $SD = 16.64$; $t(19) = 2.93$, $p = 0.009$, $d_s = 0.40$). However, the requirement for human-directed learning worsened performance satisfaction significantly ($M_{T3} = 25.25$, $SD = 16.64$; $M_{T4} = 41.00$, $SD = 23.82$; $t(19) = 4.47$, $p < 0.001$, $d_s = 0.80$). Performance satisfaction was best in the transparent learning condition (T3) and worst in the human-directed learning condition (T4).

4.4 Situation Awareness

Repeated-measures ANOVA indicate there is no significant difference in the participants' SA1 due to transparency condition, Wilks' $\Lambda = 0.931$, $F(3, 57) = 0.42$, $p = 0.742$. Throughout all conditions participants had consistently high scores on

the SA1 queries (>67%). There was a marginally significant effect of task load on SA1 scores in the opaque reasoning condition (T1), ($M_{LTL} = 0.62$, $SD = 0.19$; $M_{HTL} = 0.75$, $SD = 0.18$; $t(19) = 2.02$, $p = 0.057$, $d_s = 0.70$). When task load was high, SA1 scores were lower.

There was no significant difference in the participants' SA2 due to transparency condition, Wilks' $\Lambda = 0.813$, $F(1.87, 35.6) = 1.30$, $p = 0.306$. There was a consistent effect of task load on SA2 scores. In all conditions, when task load was high, SA2 scores were lower (T1 $d_s = 1.66$, T2 $d_s = 0.40$, T3 $d_s = 0.78$, T4 $d_s = 0.73$). This effect was most pronounced when reasoning transparency was opaque (T1: $M_{LTL} < M_{HTL}$; $M_{LTL} = 0.41$, $SD = 0.23$; $M_{HTL} = 0.09$, $SD = 0.15$, $t(19) = 5.38$, $p < 0.001$).

See Appendix L, Tables L-7 through L-9, for descriptive statistics and t-test results.

4.5 Perceived Reliability

Repeated-measures ANOVA revealed there was not a significant difference in participants' perception of agent reliability due to transparency condition, Wilks' $\Lambda = 0.459$, $F(1.88, 35.79) = 1.22$, $p = 0.304$. Paired t-tests indicate there was a significant difference in perceived reliability when human-directed learning was in effect, $t(19) = 2.56$, $p = 0.019$, $d_s = 0.45$. The agent was generally, if only slightly, perceived as being more reliable as transparency increased; however, perception of agent reliability worsened significantly in the human-directed learning condition (Fig. 10).

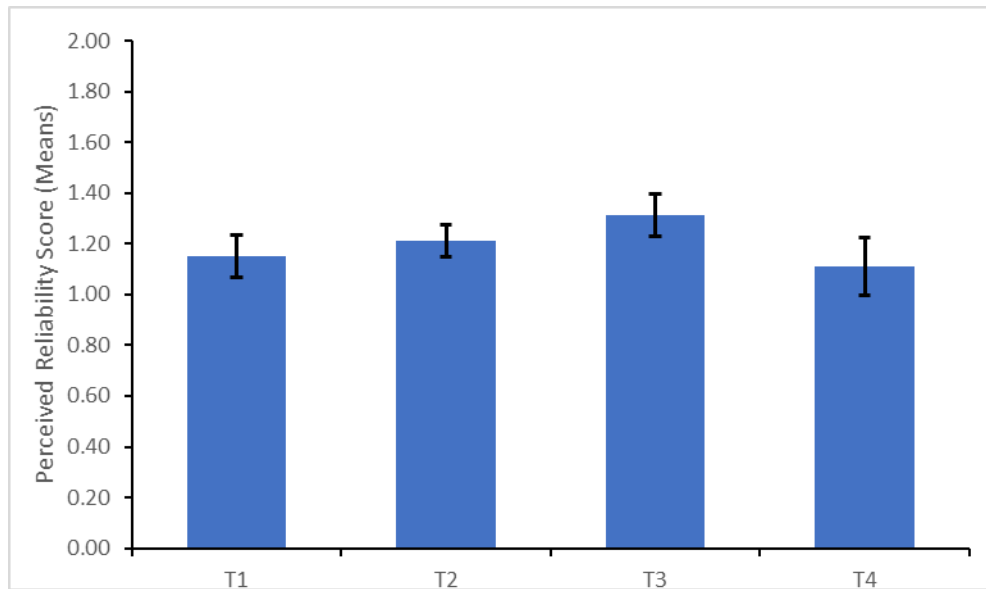


Fig. 10 Perceived reliability scores by transparency condition. Higher scores indicate higher perceived reliability. Bars denote SE.

4.6 Trust

A one-way repeated-measures ANOVA was conducted to compare scores on the TIA Scale for each of the trials (i.e., T1, T2, T3, and T4) to examine the effect of reasoning and learning transparency on trust. The means and standard deviations are presented in Appendix L, Tables L-10 and L-11. There was an overall significant effect of transparency on trust, Wilks' $\Lambda = 0.641$, $F(3, 17) = 3.18$, $p = 0.051$ (Fig. 11).

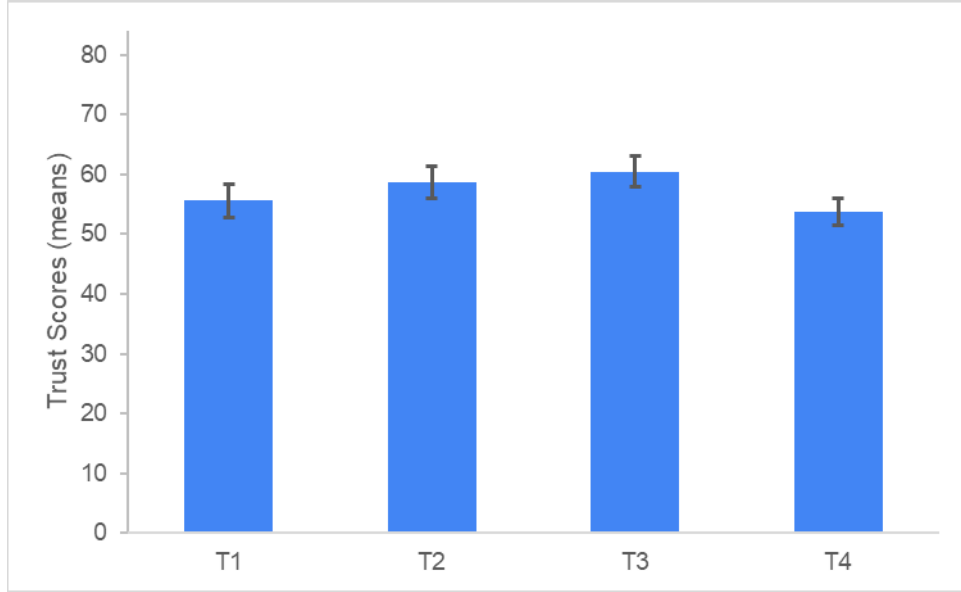


Fig. 11 Jian TIA scale scores across trials

Paired t-tests showed that trust was consistent in trials T1, T2, and T3, indicating there was no significant difference in trust scores due to either reasoning or learning transparency when comparing T1–T2 and T2–T3; however, the difference between T1 and T3 is marginally significant, showing a slight increase in trust with increased learning transparency, $M_{T1} = 55.61$, $SD = 12.68$; $M_{T3} = 60.45$, $SD = 11.51$, $t(19) = 2.04$, $p = 0.056$, $d_s = 0.40$. There was also a significant difference in trust scores due to human-directed learning. Trust scores in T4 were significantly lower than those in T3 (transparent learning) ($M_{T3} = 60.45$, $SD = 11.51$; $M_{T4} = 53.75$, $SD = 10.21$, $t(19) = 2.90$, $p = 0.019$, $d_s = 0.62$). Overall reported trust was highest when learning transparency was explicit, and lowest in the human-directed learning condition (see Appendix L, Tables L-10 and L-11).

4.7 Godspeed

Repeated-measures ANOVA indicate there is a significant difference in the participants' anthropomorphism of the agent due to transparency condition, Wilks' $\Lambda = 0.621$, $F(3, 17) = 3.46$, $p = 0.040$. Paired t-tests show that there was no

significant effect of transparency on anthropomorphism of the agent among the planned comparisons. However, there was a significant difference in anthropomorphism between T2 (transparent reasoning/opaque learning) and T4 (transparent reasoning/human-directed learning) conditions, $M_{T2} = 2.92$, $SD = 0.80$, $M_{T4} = 2.54$, $SD = 0.69$, $t(19) = 3.36$, $p = 0.003$, $d_s = 0.51$. The highest anthropomorphism scores were in T2 and the lowest were in T4 (Appendix L, Tables L-10 and L-11). See Fig. 12 for Godspeed scores by condition.

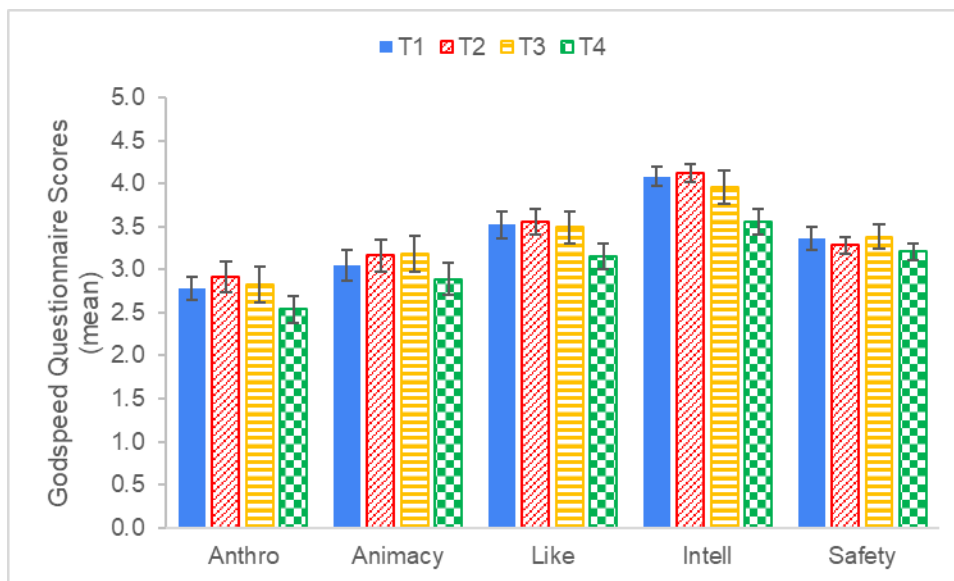


Fig. 12 Godspeed questionnaire scores by condition

There is no significant difference in the participants' perceived animacy of the agent due to transparency condition, Wilks' $\Lambda = 0.758$, $F(2.06, 17) = 1.61$, $p = 0.212$. Paired t-tests show that there was no significant effect of transparency on participants' perceived animacy of the agent among the planned comparisons. However, there was a marginally significant difference in participants' perceived animacy between T2 (transparent reasoning/opaque learning) and T4 (transparent reasoning/human-directed learning) conditions, $M_{T2} = 3.17$, $SD = 0.83$, $M_{T4} = 2.89$, $SD = 0.84$, $t(19) = 2.04$, $p = 0.055$, $d_s = 0.33$. Perceived animacy scores were higher in T2 than in T4.

There is a significant difference in the likeability of the agent due to transparency condition, Wilks' $\Lambda = 0.499$, $F(3, 17) = 5.68$, $p = 0.007$. Paired t-tests show that there is no significant effect of transparency on the likeability of the agent among the planned comparisons. However, there was a significant difference in the likeability of the agent between T2 (transparent reasoning/opaque learning) and T4 (transparent reasoning/human-directed learning) conditions $M_{T2} = 3.56$, $SD = 0.66$, $M_{T4} = 3.16$, $SD = 0.67$, $t(19) = 3.46$, $p = 0.003$, $d_s = 0.60$. The highest likeability scores were in T2 and the lowest were in T4.

There is a significant difference in the perceived intelligence of the agent due to transparency condition, Wilks' $\Lambda = 0.528$, $F(2.01, 17) = 4.29$, $p = 0.021$. Paired t-tests show that there was no significant effect of transparency on the perceived intelligence of the agent among the planned comparisons. However, there was a significant difference in the perceived intelligence between T1 (opaque reasoning and learning) and T4 (transparent reasoning/human-directed learning) conditions, $M_{T1} = 4.08$, $SD = 0.50$, $M_{T4} = 3.56$, $SD = 0.65$, $t(19) = 3.23$, $p = 0.004$, $d_s = 0.90$, and between T2 (transparent reasoning/opaque learning) and T4 conditions, $M_{T2} = 4.12$, $SD = 0.47$, $t(19) = 33.95$, $p = 0.001$, $d_s = 0.99$. The perceived intelligence scores were highest in conditions T1 and T2, and lowest in T4.

Repeated-measures ANOVA indicate there was no significant difference in the participants' perceived safety of the agent due to transparency condition, Wilks' $\Lambda = 0.890$, $F(2.14, 17) = 0.85$, $p = 0.444$. Paired t-tests show that there was no significant effect of transparency on participants' perceived safety of the agent.

4.8 Individual Differences

The influence of the ID factors PAC, intolerance of uncertainty, and personality on the dependent variables was also examined. The interpretations are made using the effect size, as the sample size for each group is very small. This information is included to identify factors of potential interest for future research.

4.8.1 Perceived Attentional Control

The PAC Survey has a potential score range of 20–80 points. The scores for the participants in this study (Min = 39, Max = 60, $Mdn_{PAC} = 54.5$, $M_{PAC} = 52.0$, $SD = 6.40$, $SE = 1.43$) span the midpoint of the range of potential scores (i.e., 50). Differences due to PAC scores were evaluated using independent samples t-tests comparing high PAC scores (>54.5) to low PAC scores (≤ 54.5). Descriptive statistics and t-test results are shown in Appendix L, Tables L-12 through L-16.

4.8.1.1 Task Performance

Persons in the high PAC group had more correct identifications than those in the low PAC group in both the T2 ($M_{HPAC} = 94.82$, $SD = 3.70$, $M_{LPAC} = 92.69$, $SD = 3.88$, $d_s = 0.56$) and T4 ($M_{HPAC} = 94.55$, $SD = 3.37$, $M_{LPAC} = 92.43$, $SD = 3.99$, $d_s = 0.58$) conditions, as well as shorter response times in condition T4 for both correct ($M_{HPAC} = 12.79$, $SD = 5.16$, $M_{LPAC} = 14.99$, $SD = 4.89$, $d_s = 0.44$) and incorrect ($M_{HPAC} = 18.06$, $SD = 5.54$, $M_{LPAC} = 20.02$, $SD = 6.28$, $d_s = 0.33$) responses (Fig. 13).

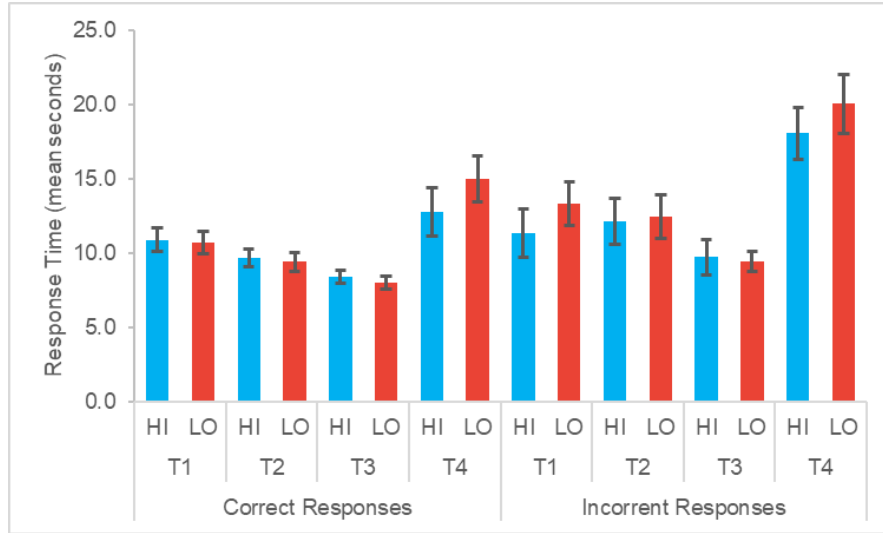


Fig. 13 High PAC vs. low PAC response times for correct and incorrect identifications

4.8.1.2 Workload

Persons in the high PAC group reported greater MD in conditions T1 ($d_s = 0.49$), T3 ($d_s = 0.53$), and T4 ($d_s = 0.98$), greater effort in conditions T1 ($d_s = 0.50$) and T4 ($d_s = 0.61$), and greater physical demand (T1 $d_s = 0.50$, T2 $d_s = 0.58$, T3 $d_s = 0.78$, and T4 $d_s = 0.72$) in all conditions than those in the low PAC group (Fig. 14). Persons in the low PAC group reported greater TD demand in the T1 ($d_s = 0.50$) condition and greater performance satisfaction in condition T4 ($d_s = 0.53$) than those in the high PAC group.

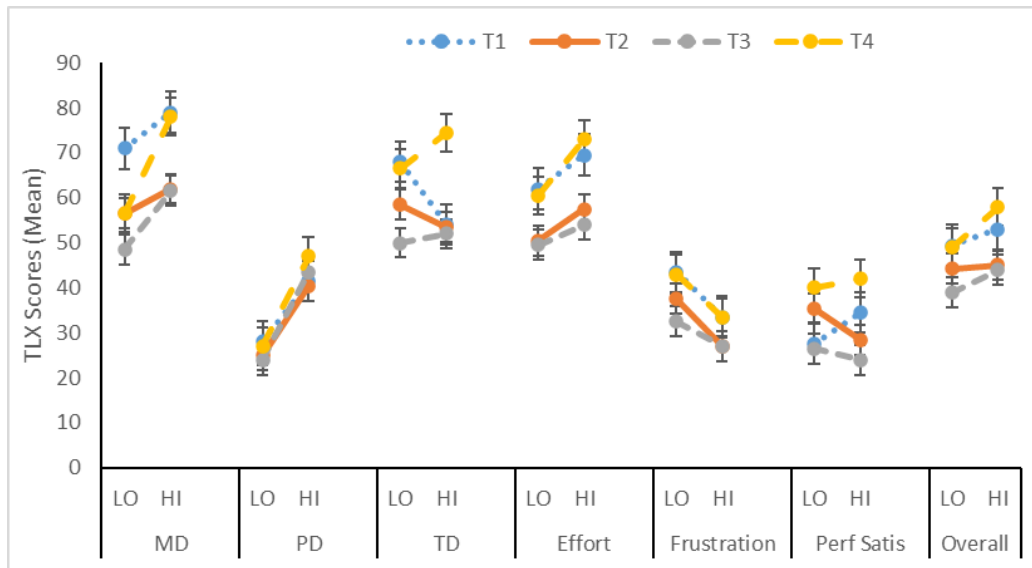


Fig. 14 NASA-TLX scores by low vs. high PAC

4.8.1.3 Situation Awareness

Persons in the high PAC group had higher SA scores in conditions T2, T3, and T4 than those in the low PAC group for both SA1 (T2 $d_s = 0.72$, T3 $d_s = 0.65$, and T4 $d_s = 1.04$) and SA2 (T1 $d_s = 0.74$, T2 $d_s = 0.48$, T3 $d_s = 0.49$, and T4 $d_s = 0.85$) queries. In conditions T1, T2 and T4, the high PAC group perceived the robot as more reliable than those in the low PAC group (T1 $d_s = 0.54$, T2 $d_s = 0.44$, and T4 $d_s = 0.77$).

4.8.1.4 Trust

Persons in the high PAC group reported greater trust in the agent than those in the low PAC group in all conditions, notably in scores in conditions T2 $d_s = 0.54$, T3 $d_s = 0.44$, and T4 $d_s = 1.21$ (Fig. 15).

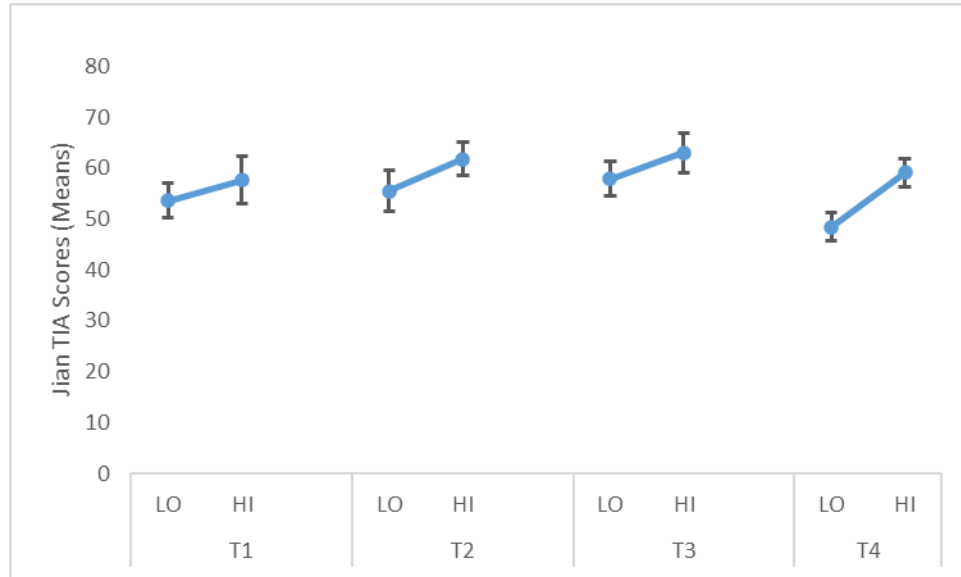


Fig. 15 Jian TIA scale scores by low vs. high PAC

4.8.1.5 Godspeed

There were no differences due to PAC in participant perceptions of the agent in the opaque (T1) condition for any factor, and no difference in perception of animacy for any condition. As transparency increased, the differences due to PAC became apparent. High PAC individuals perceived the agent to be more likeable (T2 $d_s = 0.28$, T3 $d_s = 0.87$, and T4 $d_s = 0.98$), more intelligent (T2 $d_s = 0.62$, and T4 $d_s = 1.11$), and safer (T2 $d_s = 1.79$, T3 $d_s = -1.26$, and T4 $d_s = 0.87$) than the low PAC group.

4.8.2 Uncertainty

The IUS questionnaire assesses two factors related to how an individual responds to uncertainty: PA and IA. The PA score has a potential range of 7–28 points, IA score has a potential range of 5–20 points. The scores for PA in this study, Min = 10, Max = 24, $Mdn_{PA} = 17.0$, $M_{PA} = 16.6$, $SD = 4.02$, $SE = 0.90$, span the midpoint of the range of potential scores (i.e., $PA = 17.5$). However, the scores for IA, Min = 5, Max = 13, $Mdn_{IA} = 8.0$, $M_{IA} = 8.0$, $SD = 2.36$, $SE = 0.53$, are nearly all below the original scales' design midpoint (i.e., $IA = 12.5$), therefore all participants could be considered low IA in relation to the theoretical midpoint. As such, IA was examined to see if even incremental differences in that factor led to differential results. Differences due to IUS scores were evaluated using independent samples t-tests comparing high scores ($PA \geq 17.0$, $IA \geq 8.0$) to low scores ($PA < 17.0$, $IA < 8.0$). Descriptive statistics and t-test results are shown in Appendix L, Tables L-17 through L-26.

4.8.2.1 Task Performance

Persons in the high PA group had fewer correct identifications than those in the low PA group in the T1 ($d_s = 0.62$), T2 ($d_s = 0.43$), and T4 ($d_s = 0.90$) conditions (Fig. 16). The high PA individuals had faster response times in condition T2 ($d_s = 0.60$); however, they had slower response times in condition T4 both when responding correctly ($d_s = 0.60$) and incorrectly ($d_s = 0.73$) than the low PA individuals.

Persons with higher IA had fewer correct identifications than those with lower IA in the T2 ($d_s = 0.86$) condition. The higher IA individuals had faster response times when responding correctly in conditions T2 ($d_s = 0.64$) and T3 ($d_s = 0.69$), as well as when responding incorrectly in condition T3 ($d_s = 0.67$). However, they had slower response times in condition T1 when responding incorrectly ($d_s = 0.56$) than the lower IA individuals (Fig. 17).

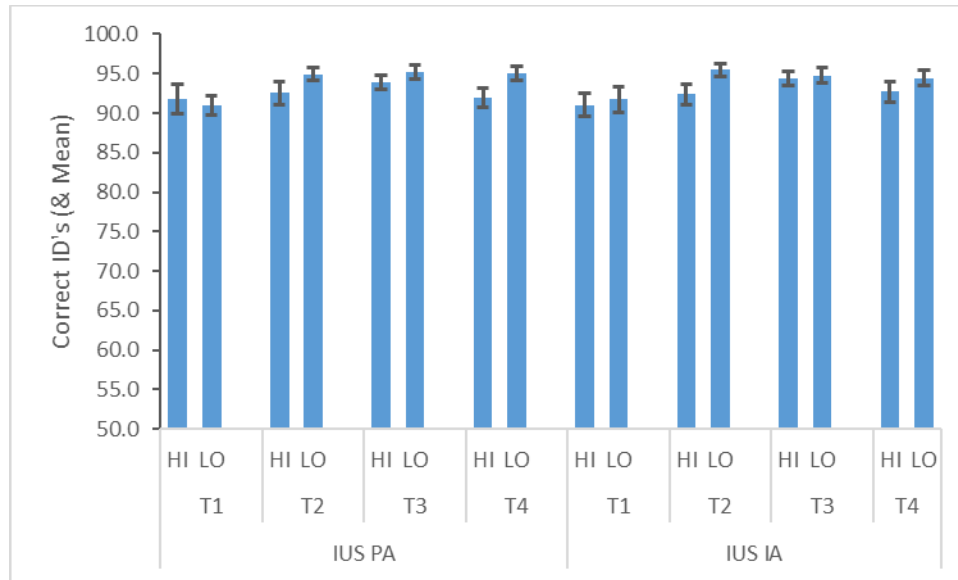


Fig. 16 Task performance (correct IDs) by prospective anxiety and inhibitory anxiety

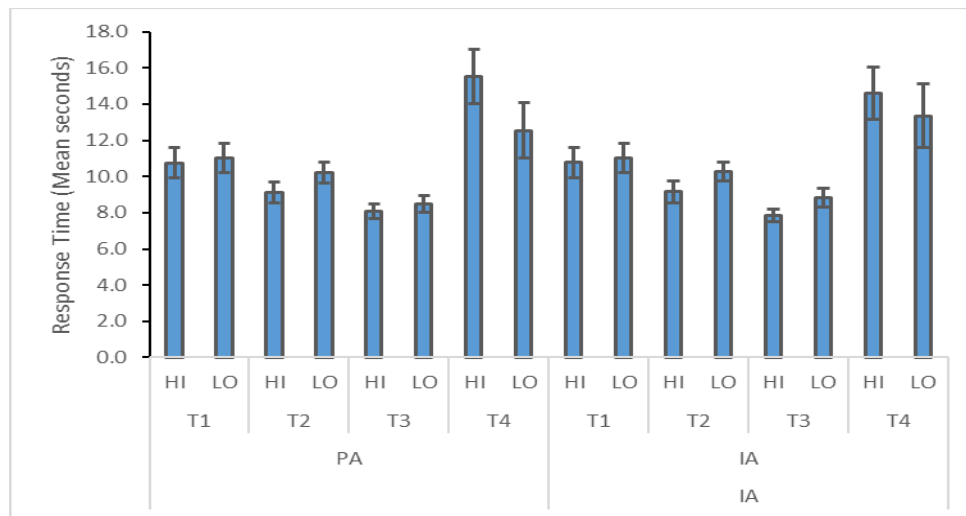


Fig. 17 Response times by high/low prospective and inhibitory anxiety

4.8.2.2 Workload

Persons in the high PA group reported lower MD than those in the low PA group in all conditions; T1 ($d_s = 0.70$), T2 ($d_s = 0.53$), T3 ($d_s = 0.36$) and T4 ($d_s = 0.44$). They also reported lower effort in the T1 ($d_s = 0.64$) and T2 ($d_s = 0.50$) conditions than those in the low PA group, however they reported greater frustration than the low PA group in condition T4 ($d_s = 0.68$). Persons in the high PA group reported greater performance satisfaction than the low PA group in all conditions, particularly T1($d_s = 0.69$) and T4 ($d_s = 0.51$). There were no noticeable differences between the two groups in reported PDI or TD (Fig. 18).

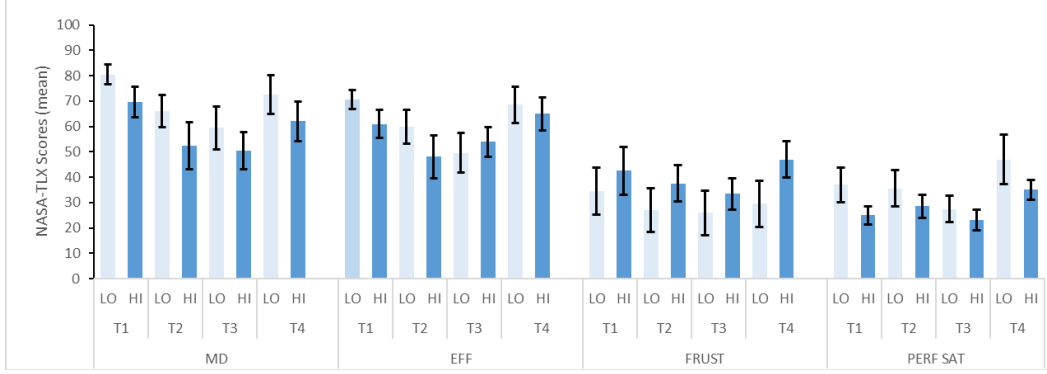


Fig. 18 Selected NASA-TLX scores by high/low prospective anxiety. MD = mental demand, EFF = effort, FRUST = frustration, PERF SAT = performance satisfaction. Bars denote SE.

Persons in the high IA group reported lower MD than those in the low IA group in conditions T1 ($d_s = 0.70$) and T2 ($d_s = 0.45$), lower TD in all conditions; T1 ($d_s = 1.07$), T2 ($d_s = 0.97$), T3 ($d_s = 0.52$) and T4 ($d_s = 0.37$), and greater PD in T1 ($d_s = 1.54$). However, they reported worse performance satisfaction than the low IA group in conditions T1 ($d_s = 0.69$) and T4 ($d_s = 0.50$). There were no noticeable differences between the two groups in reported Effort or Frustration (Fig. 19).

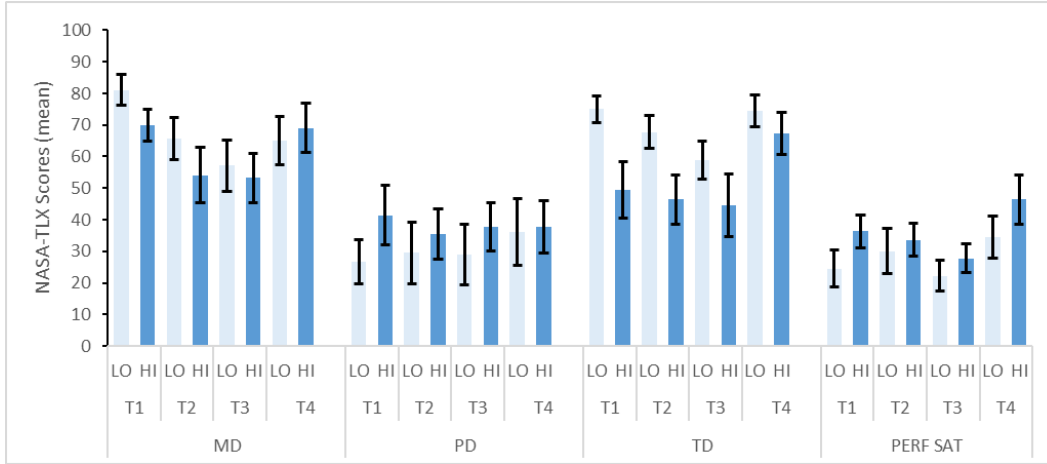


Fig. 19 Selected NASA-TLX scores by high/low inhibitory anxiety. MD - mental demand, EFF = effort, FRUST = frustration, PERF SAT = performance satisfaction. Bars denote SE.

4.8.2.3 Situation Awareness

There were few differences in SA scores due to PA, and most had medium effect size. Individuals with low PA scored higher on SA1 queries (T1 $d_s = 0.45$ and T2 $d_s = 0.71$), SA2 queries in T4 ($d_s = 0.40$) and perceived the agent to be more reliable (T2 $d_s = 0.44$ and T4 $d_s = 0.55$) than individuals in the high PA group.

There were few differences in SA scores due to IA. Individuals with low IA scored higher on SA1 queries (T2 $d_s = 0.82$), T3 $d_s = 0.47$, and T4 $d_s = 0.50$), higher on SA2 queries in T2 ($d_s = 0.42$), and more reliable in T2 ($d_s = 0.61$) than individuals

in the high PA group. However, in T3 high IA individuals perceived the agent to be more reliable than low IA individuals ($d_s = 0.43$), although this was a small-medium effect size.

4.8.2.4 Trust

Participants in the low PA group reported greater trust in the agent in conditions T2 ($d_s = 0.51$), T3 ($d_s = 0.50$) and T4 ($d_s = 1.03$) than individuals in the high PA group. Participants in the low IA group reported greater trust in the agent in all conditions, notably conditions T1 ($d_s = 0.86$) and T4 ($d_s = 1.04$), than individuals in the high IA group.

4.8.2.5 Godspeed

Low PA individuals perceived the agent as more likeable, T2 ($d_s = 1.39$), T3 ($d_s = 0.93$) and T4 ($d_s = 0.90$), more intelligent, T2 ($d_s = 0.82$) and T4 ($d_s = 0.56$) and safer, T2 ($d_s = 0.74$), T3 ($d_s = 1.09$) and T4 ($d_s = 0.87$) than individuals in the high PA group (Fig. 20). Low IA individuals also perceived the agent as more likeable, T1 ($d_s = 0.62$), T2 ($d_s = 1.08$), T3 ($d_s = 0.62$) and T4 ($d_s = 0.82$), more intelligent, T1 ($d_s = 0.80$), T2 ($d_s = 0.68$) and T4 ($d_s = 1.54$) and Safer, T1 ($d_s = 0.57$), T2 ($d_s = 1.42$), T3 ($d_s = 0.98$) and T4 ($d_s = 1.26$) than individuals in the high IA group (Fig. 21). There were no differences in anthropomorphism or animacy scores due to either PA or IA.

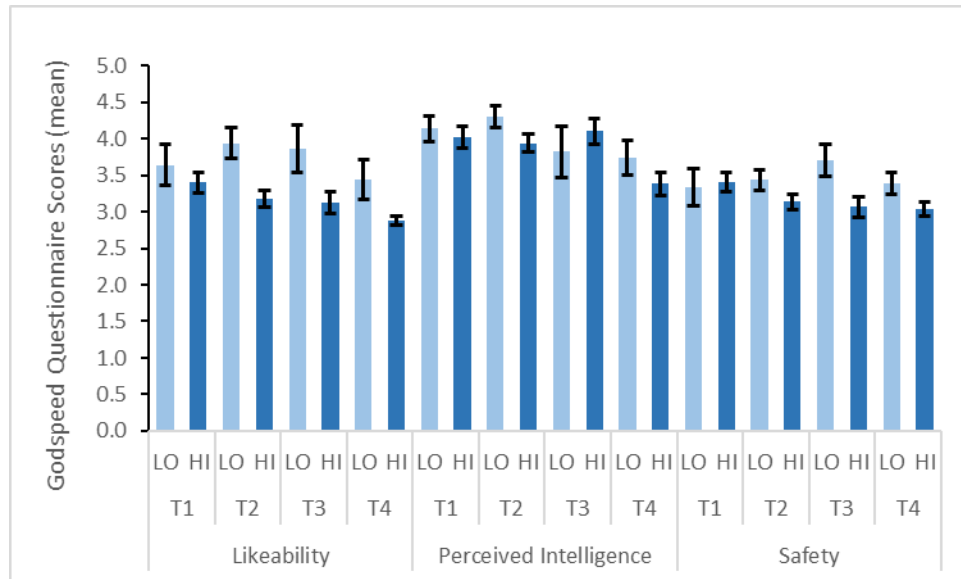


Fig. 20 Selected Godspeed questionnaire scores by high/low prospective anxiety. Bars denote SE.

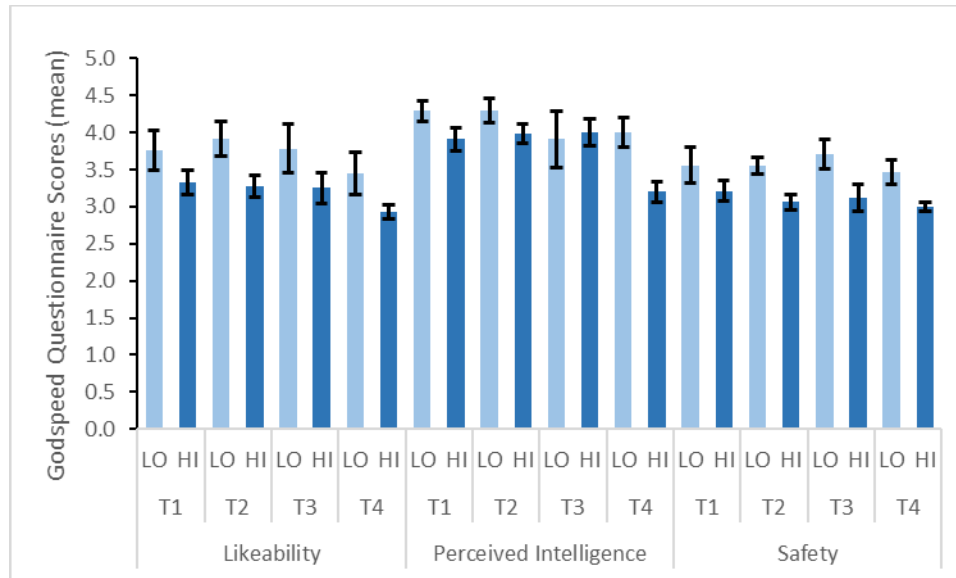


Fig. 21 Selected Godspeed questionnaire scores by high/low inhibitory anxiety. Bars denote SE.

4.8.3 Personality

The mini-IPIP survey was used to assess participant's personality on five factors. Each factor has a potential score range of 4–20 points, midpoint = 12. The scoring data for the participants in this study are shown in Table 2.

Table 1 Descriptive statistics for personality factor data

Factor	<i>M</i>	<i>Mdn</i>	<i>SD</i>	Min	Mid	Max	N _{HI}	N _{LO}
IPIP-E	12.35	13.0	3.65	5.0	11.00	17.0	11	9
IPIP-A	16.45	16.0	2.13	12.0	16.00	20.0	9	11
IPIP-C	11.90	12.0	2.91	6.0	11.50	17.0	12	8
IPIP-N	11.90	11.0	4.35	4.0	12.00	20.0	10	10
IPIP-O	17.05	17.0	1.91	13.0	16.50	20.0	13	7

The scores for factors E(*xtroversion*), C(*onscientiousness*), and N(*euroticism*) span the midpoint of the range of potential scores (i.e., 12), so in those cases the differences examined will be due to low or high rating on that factor. The scores for factors A(*greeableness*) and O(*penness*) are all above the theoretical midpoint (i.e., 11 points) of the measure, as such, all participants were in the high category; however, they will still be examined to see if even incremental differences in that factor led to differential results. See Fig. 22 for high/low group membership definitions. Differences due to factor scores were evaluated using independent samples t-tests. Descriptive statistics and t-test results are shown in Appendix L, Tables L-27 through L-51.

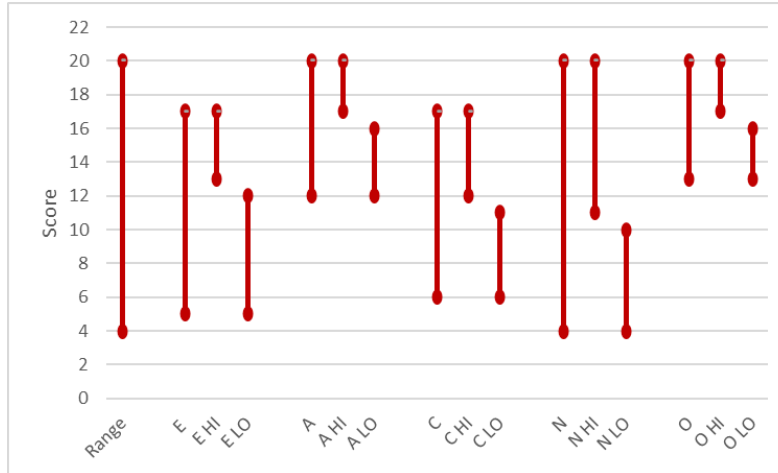


Fig. 22 Range of scores for each personality factor. The measure's potential range of scores is shown as the left-most entry. Each personality factor's scores are shown as overall, then how they were sorted into high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.

4.8.3.1 Task Performance

Extroversion. Persons in the high extroversion (E) group had fewer correct identifications than those in the low E group in the T1 ($d_s = 0.72$) and T2 ($d_s = 0.56$) conditions. The high E individuals had slower response times in conditions T4 ($d_s = 0.80$), and in condition T3 ($d_s = 1.01$) when answering correctly, than the low E individuals.

Agreeableness. Persons in the high agreeableness (A) group had more correct identifications than those in the low A group in the T2 ($d_s = 0.70$) and T4 ($d_s = 0.57$) conditions. The high A individuals had faster response times in condition T4 ($d_s = 1.10$), and in condition T3 ($d_s = 0.56$) when answering correctly, than the low A individuals.

Conscientiousness. There was no difference in performance due to Conscientiousness (C) in any condition.

Neuroticism. There was no difference in the percentage of correct identifications due to neuroticism (N). Persons in the high N group had faster response times in condition T3 ($d_s = 0.74$) for correct IDs; however they had slower response times in condition T2 ($d_s = 0.67$) when answering incorrectly than the low N individuals.

Openness. Persons in the high openness (O) group had fewer correct identifications than those in the low O group in the T1 ($d_s = 0.53$) condition. The high O individuals had faster response times in conditions T2 ($d_s = 0.57$) and T3 ($d_s = 0.69$) than the low O individuals, and in condition T4 ($d_s = 0.71$) when answering incorrectly (see Table 3).

Table 3 Task performance high/low group outcomes at a glance. X indicates high group had better performance; O indicates high group had worse performance. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.

Performance measure	Condition	E	A	C	N	O
Correct IDs	T1	O				O
	T2	O	X			
	T3					
	T4		X			
Response Time	T1					
	T2				O	X
	T3	O	X		X	X
	T4	O	X			X

4.8.3.2 Workload

Extroversion. In condition T1, participants in the high E group reported lower MD ($d_s = 0.70$), yet greater PD ($d_s = 0.54$) than their low E counterparts. High E participants reported lower TD than the low E group in all conditions; T1 ($d_s = 1.07$), T2 ($d_s = 0.97$), T3 ($d_s = 0.52$) and T4 ($d_s = 0.37$). However, low E participants reported greater performance satisfaction, T1 ($d_s = 0.69$), T2 ($d_s = 0.50$), than the low E group. There were no differences in reported scores for effort or frustration factors.

Agreeableness. High A participants reported lower PD than the low As individuals in all conditions; T1 ($d_s = 0.99$), T2 ($d_s = 2.37$), T3 ($d_s = 1.86$) and T4 ($d_s = 1.44$). Low A individuals reported lower TD than the high A participants in all conditions, with a medium effect size in T3 ($d_s = 0.49$). High A participants reported lower effort in T1 ($d_s = 0.49$), lower frustration in T4 ($d_s = 0.66$), and greater performance satisfaction in T4 ($d_s = 0.55$) than the low A group. There was no reported difference in MD between the two groups.

Conscientiousness. There were few differences between high and low C groups in reported workload factors. In condition T4, low C individual reported lower MD ($d_s = 0.55$) than the high C group. High C individuals reported lower TD in condition T1 ($d_s = 0.53$), and greater performance satisfaction in conditions T2 ($d_s = 0.74$) and T3 ($d_s = 0.63$) than the low C participants. There were no differences in reported scores for the PD, effort, or frustration factors.

Neuroticism. Participants in the high N group reported greater TD in conditions T1 ($d_s = 0.66$) and T2 ($d_s = 0.59$), more effort in condition T1 ($d_s = 0.64$), and greater frustration in condition T4 ($d_s = 0.91$) than the low N participants. However, they also reported greater performance satisfaction in all conditions, particularly T3

($d_s = 0.52$) and T4 ($d_s = 0.42$). There were no differences in reported scores for the MD or PD factors.

Openness. Participants in the high O group reported greater MD in conditions T1 ($d_s = 0.62$) and T2 ($d_s = 0.76$), lower PD in all conditions T1 ($d_s = 0.67$), T2 ($d_s = 1.03$), T3 ($d_s = 0.87$) and T4 ($d_s = 0.86$), more effort in condition T1 ($d_s = 0.59$), and less frustration in conditions T2 ($d_s = 0.67$), T3 ($d_s = 0.53$) and T4 ($d_s = 0.85$) than the low O participants. There were no differences in reported scores for the TD or performance satisfaction factors (see Table 4).

Table 4 NASA-TLX high/low group outcomes at a glance. X indicates high group had lower scores; O indicates high group had higher scores. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.

Workload measure	Condition	E	A	C	N	O
Mental Demand	T1	X				O
	T2					O
	T3					
	T4			O		
Temporal Demand	T1	X	O	X	O	
	T2	X	O		O	
	T3	X	O			
	T4	X	O			
Physical Demand	T1	O	X			X
	T2		X			X
	T3		X			X
	T4		X			X
Effort	T1		X		O	O
	T2					
	T3					
	T4					
Frustration	T1					
	T2					X
	T3					X
	T4		X		O	X
Performance Satisfaction	T1	X			X	
	T2	X		X	X	
	T3			X	X	
	T4		X		X	

4.8.3.3 Situation Awareness

Extroversion. In condition T4, participants in the high E group had higher SA Level 1 scores ($d_s = 0.44$) than the low E group. There were no differences in SA Level 2 scores due to E. Participants in the low E group perceived the agent to be more reliable in conditions T2 ($d_s = 0.72$), T3 ($d_s = 0.41$), and T4 ($d_s = 0.47$) than their high E counterparts.

Agreeableness. High A participants had higher SA Level 1 scores in conditions T1 ($d_s = 0.56$), T2 ($d_s = 0.53$) and T3 ($d_s = 0.76$), higher SA Level 2 scores in conditions T1 ($d_s = 0.55$) and T3 ($d_s = 0.63$), and perceived the agent to be more reliable in conditions T1 ($d_s = 0.49$) and T3 ($d_s = 0.98$) than the low A group.

Conscientiousness. The high C individuals had higher SA Level 1 scores in conditions T1 ($d_s = 0.45$), T2 ($d_s = 0.82$), T2 ($d_s = 0.61$), and T3 ($d_s = 0.50$), higher SA Level 2 scores in condition T4 ($d_s = 0.62$), and higher perceived reliability in condition T4 ($d_s = 0.60$) than the low C group.

Neuroticism. Participants in the low N group reported higher SA Level 1 scores in conditions T2 ($d_s = 0.44$) and T3 ($d_s = 0.51$), higher SA Level 2 scores in condition T4 ($d_s = 0.54$) and higher perceived reliability of the agent in condition T4 ($d_s = 0.44$) than the high N individuals.

Openness. There were few differences in SA scores due to O. Participants in the high O group had higher SA Level 1 scores in condition T1 ($d_s = 0.80$), higher SA Level 2 scores in condition T1 ($d_s = 0.74$), and higher perceived reliability of the agent in condition T1 ($d_s = 0.54$) than the low O individuals (see Table 5).

Table 5 Sagat-Probe high/low group outcomes at a glance. X indicates high group had higher scores; O indicates high group had lower scores. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.

SA measure	Condition	E	A	C	N	O
SA1	T1		X	X		X
	T2		X	X	O	
	T3		X	X	O	
	T4	X		X		
SA2	T1		X			X
	T2					
	T3		X			
	T4			X	O	
Perceived Reliability	T1		X			X
	T2	O				
	T3	O	X			
	T4	O		X	O	

4.8.3.4 Trust

There were no notable differences in reported trust in the agent due to either extroversion or agreeableness. Greater trust in the agent was reported by high C individuals in conditions T2 ($d_s = 1.10$), T3 ($d_s = 0.77$) and T4 ($d_s = 1.19$), low N individuals in condition T4 ($d_s = 0.70$), and high O individuals in condition T4 ($d_s = 0.67$) than their respective counterparts (see Table 6).

Table 6 Jian TIA scale high/low group outcomes at a glance. X indicates high group had higher scores; O indicates high group had lower scores. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.

Jian TIA measure	Condition	E	A	C	N	O
Trust	T1					
	T2			X		
	T3			X		
	T4			X	O	X

4.8.3.5 Godspeed

Extroversion. There were no differences in perceived likeability, intelligence, or safety due to E. High E individuals anthropomorphized the agent more in condition T3 ($d_s = 0.49$) and perceived the agent to have greater animacy in conditions T1 ($d_s = 0.54$) and T2 ($d_s = 0.58$) than the low E individuals.

Agreeableness. There were no differences in perceived likeability due to agreeableness. High A individuals anthropomorphized the agent more in condition T4 ($d_s = 0.64$), perceived the agent to have greater animacy in conditions T2 ($d_s = 0.72$) and T4 ($d_s = 0.70$), greater intelligence in condition T2 ($d_s = 0.49$), and safer in condition T4 ($d_s = 0.52$) than their low A individuals.

Conscientiousness. There were numerous and consistent differences in perceptions of the agent due to C. In all conditions, high C individuals anthropomorphized the agent more, T1 ($d_s = 0.88$), T2 ($d_s = 0.82$), T3 ($d_s = 0.98$) and T4 ($d_s = 0.59$), and perceived the agent to have greater animacy, T1 ($d_s = 1.29$), T2 ($d_s = 1.14$), T3 ($d_s = 0.67$) and T4 ($d_s = 0.58$), than their low C counterparts. They also perceived the agent to be more likeable in conditions T2 ($d_s = 0.91$) and T3 ($d_s = 0.74$), more intelligent in conditions T1 ($d_s = 0.61$), T2 ($d_s = 0.62$), and T4 ($d_s = 0.61$), and safer in condition T2 ($d_s = 0.83$) than the low C individuals.

Neuroticism. There were few notable differences in perceptions of the agent due to N. When agent reasoning was opaque (condition T1), High N individuals anthropomorphized the agent more ($d_s = 0.82$), perceived the agent to have greater animacy ($d_s = -0.50$), to be more intelligent ($d_s = 0.48$), and to be safer ($d_s = 0.67$) than the low N individuals. However, low N individuals perceived the agent to be more intelligent in the human-directed learning condition (T4; $d_s = 0.63$) than the high N individuals. There were no differences in perceived likeability due to N.

Openness. There were no differences in anthropomorphism, perceived animacy or intelligence due to O. High O individuals reported the agent to be more likeable

(T2: $d_s = 0.81$, T4: $d_s = 0.72$) and safer (T4: $d_s = 0.78$) than the low O individuals (see Table 7).

Table 7 Godspeed questionnaire high/low group survey outcomes at a glance. X indicates high group had higher scores; O indicates high group had lower scores. Shaded cells had no difference due to high/low group membership. E = extroversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.

Godspeed measure	Condition	E	A	C	N	O
Anthropomorphism	T1			X	X	
	T2			X		
	T3	X		X		
	T4		X	X		
Animacy	T1	X		X	X	
	T2	X	X	X		
	T3			X		
	T4		X	X		
Likeability	T1					
	T2			X		X
	T3			X		
	T4					X
Perceived Intelligence	T1			X	X	
	T2		X			
	T3					
	T4			X	O	
Perceived Safety	T1				X	
	T2			X		
	T3					
	T4		X			X

5. Discussion

The purpose of this research was to examine human-agent teaming with an LCA, specifically the effects of transparent agent reasoning and learning on the human teammate. Participants teamed with a simulated AiTR agent to patrol within a simulated urban environment and identify potential threats when encountered. Occasionally the AiTR could not complete a classification and the participant had to complete those, with the knowledge that the agent would later use this provided information to update and refine its programming. Participants completed four trials, one at each transparency level. In the highest-level trial (i.e., human-directed learning) the participant directly updated the agent reasoning and could change the agent reasoning on factors they disagreed with.

As expected, transparent agent reasoning improved human task performance by increasing the percentage of overall correct classifications, as well as reduced response times. The addition of agent reasoning transparency reduced the number of incorrect rejections of the agent's classifications. However, neither transparent

learning nor human-directed learning improved the number of correct classifications completed. Transparent agent reasoning improved (reduced) response times, and transparent agent learning further reduced response times. These findings support prior research findings that, in general, increasing transparency improves task response time, even though it may not always result in improved task performance (Wright et al. 2017, 2020a; Bhaskara et al. 2021). Human-directed learning resulted in the longest response times of any condition. High versus low task load evaluations were as expected with fewer correct classifications in the HTL segments in all conditions. However, HTL increased response time only in the human-directed learning condition. These findings are strong evidence that human-directed learning should not be conducted during time critical missions.

Workload was expected to decrease as transparency increased, and that is largely what the results indicate. Across all workload factors, the explicit learning condition (T3) had the lowest workload factor scores and the greatest reported performance satisfaction. The opaque reasoning (T1) and the human-directed learning (T4) conditions were expected to elicit the highest workload scores, which is what was found, although there were a few unexpected outcomes. The opaque condition elicited the greatest MD. High MD in the opaque condition is not too surprising, as providing no information has been found to elicit higher workload than even supplying erroneous information (Dadashi et al. 2013). However, MD was also higher in the opaque condition than in the human-directed learning condition, which required the participant to perform more actions and decisions than any of the other conditions.

The human-directed learning condition had significantly higher TD and the worst performance satisfaction than any other condition. It stands to reason that the additional task of teaching the agent had the participants feeling rushed (even though their response time was longer), but also having poor performance satisfaction indicates they felt their (or the agent's) task performance was insufficient or lacked confidence in their performance. For the effort and frustration workload factors, the human-directed learning condition had similar outcomes as the opaque reasoning condition, which were considerably worse than the transparent learning conditions T2 and T3.

Participants' SA was evaluated by assessing their understanding of where the vehicle was in the city or comprehension of factors affecting route divergence. This entailed periodically reviewing the vehicle location in the map, the route planning agent's comparison report of available routes, and the commanders chat messages. Increased transparency was expected to improve participants' SA by freeing up their attentional resources by offering information from the agent in a direct

method. However, human-directed learning was expected to interfere with the participant's ability to stay aware of the vehicle information, primarily because of the increase in task load in that condition. Overall, agent transparency did not affect participant's perception of elements in the environment, the vehicle's location, or their understanding why the vehicle was on the current route.

Generally, participants' perception of the agent's reliability was improved by increased transparency. There was a persistent notion that the agent was less reliable than it truly was, but this steadily improved as transparency increased. However, when the participant was tasked with teaching the agent (T4), the agent was perceived to be significantly less reliable than in the other conditions. It is possible that in the lower levels of agent learning transparency, the participant was simply not attending to how the agent was interpreting their input or what specific conditions caused the agent's inability to complete a classification. When the participant had to directly inform the agent's learning process by completing classifications (T4) this increased their awareness of the frequency of the agent's inability to complete classifications and the types of conditions that caused the agent's confusion.

Trust in the agent was not affected by either reasoning or learning transparency. In each successive trial human trust in the agent incrementally increased. However, when tasked with directly teaching the agent (T4), trust worsened significantly. This may be due to increasing the participant's awareness of the type and frequency of the agent's inability to complete classifications.

An operator's perception of an agent can influence their interactions with said agent, particularly misuse/disuse, or complacent behaviors (Parasuraman et al. 2000; Robinette et al. 2016; Pak et al. 2020). In the current study, there was little to no difference in participant perception of the agent's anthropomorphism, animacy, likeability, intelligence, or safety due to transparency of reasoning or learning. However, there were consistent and significant differences in perceived anthropomorphism, animacy, likeability, and intelligence between the human-directed learning condition and the other conditions. When tasked with teaching the agent, the participant viewed the agent as less anthropomorphic, animate, likeable, and intelligent. These more negative assessments, particularly when combined with the perception of reduced reliability, could create a situation wherein the participant engages in over-monitoring behavior (Hinds et al. 2004).

The influence of several individual difference factors (i.e., attention control, response to uncertainty, personality) on participant performance and perceptions of the agent were also examined. Greater attention control was extremely beneficial in terms of task performance, trust in the system, SA, and perceptions of the agent.

However, when the individual workload factors were examined, higher attention control individuals reported greater workload for mental and physical demand than those reporting lower attention control. Those reporting lower attention control reported greater TD, indicating they felt rushed; however, they also reported the greatest performance satisfaction when they were tasked with teaching the agent.

Anxiety about uncertainty, both prospective and inhibitory, had a largely detrimental effect on participant performance. Higher reported anxiety resulted in lower task performance, SA, trust in the agent, and worse perceptions of the agent's likeability, intelligence, and safety. However, high anxiety may have had a buffering effect on perceived workload. High PA individuals reported lower MD in all conditions, and high IA individuals reported lower MD when reasoning and learning was opaque. High PA individuals reported lower effort, yet greater frustration, and greater performance satisfaction. High IA individuals reported lower TD and lower performance satisfaction. It appears that, for the high anxiety individuals, anticipation of the mission interfered with proper trust calibration, task performance, and negatively influenced interactions with the agent, yet inoculated the individual against high workload on several factors.

Evaluations of individual differences in the variables of interest (performance, workload, etc.) due to personality showed minor and varied outcomes. Personality factors appeared to have little influence on participant task performance, although those high in A and O had faster response times when transparency increased. Perceived workload differences due to personality were varied; high E individuals reported lower temporal demand in all conditions, while high A individuals reported greater TD. Both high A and O individuals reported lower physical demand in all conditions, and the high O individuals also reported lower frustration in the increased transparency conditions. High N individuals reported better performance satisfaction in all conditions than their low N counterparts. High C individuals had higher SA1 scores in conditions T1, T2, and T3, and higher SA2 scores in T3 and T4 than their low C counterparts. High E individuals perceived the agent as less reliable in all conditions than the low E group. Personality factors appeared to have little effect on participant trust in the agent, although high C individuals reported higher trust than their low C counterparts when transparency was increased (T2, T3, and T4). There was also little effect on the Godspeed factors due to personality differences, except for high C individuals, who appeared to anthropomorphize the agent more and attribute greater animacy to the agent in all conditions than the low C individuals, as well as influence differences in perceived likeability and intelligence due to transparency level. Taken together, these results indicate that attempting to control for differences in personality when evaluating outcomes in human-agent teaming may not be particularly helpful, as their

influence appears to be sporadic and inconsistent. For example, Cs (considered to be the trait associated with a sense of duty and goal-oriented behavior, which would contribute to better task performance) and N (associated with poor self-regulation and difficulty handling stress, which would contribute to worse task performance), had no consistent influence on either task performance or perceived workload, which would appear to be contrary to accepted norms (Szalma and Taylor 2011; Servaas et al. 2013; Lin et al. 2020).

5.1 Limitations

There are several notable limitations in this study.

It is indistinguishable whether in T4 the participant perception of agent reliability was influenced by forced attention to the agent's shortcomings or by increased workload. In T3 the participant could have simply not been attending to the agent's reported use of their inputs, whereas in T4 the participant had to attend closely to the individual factors used in classifications. This appears to be an important distinction to tease apart in future studies.

The assignment in this study was not task load-intensive, and the participant was primarily focused on a single task, rather than balancing several tasks. These evaluations should be replicated in a higher demand setting to better understand how increased stress influences the outcomes.

5.2 Future Directions

In this study the agent was perfectly reliable, although this was not explicitly shared with the participant. It is important to understand how teaming with an agent that is less than perfectly reliable differs from these findings, as it is unlikely any fielded agent will be 100% reliable in the field, primarily due to dynamic, evolving environments. Additionally, assessing trust as a continuous, dynamic state rather than a single, subjective outcome should better illustrate how intervening to teach the agent, rather than only supervising the agent's actions, affects the participant's evaluations of the agent in real time.

6. Conclusion

Working with agents that continue to learn post-deployment is a future capability that is fast approaching. Understanding how teaming with these LCAs affects the human teammates is crucial to the successful deployment of these systems. As demonstrated in previous studies, transparency of the agent's reasoning supports team performance in several important ways: improved task performance, reduced

response times, reduced MD, and greater SA. This study demonstrated that it is equally important to be transparent as to how the agent is using supplied information to update its algorithms and alter its future behavior. Transparency of agent learning resulted in improved response times, reduced overall workload, and increased performance satisfaction. Understanding when the agent could not complete a classification or witnessing how the agent inferred information from the participant input did not negatively influence the participant's perception of agent reliability, in fact perceived reliability was significantly higher when this information was available. However, directly teaching the agent had a decidedly negative impact on response times, perceived workload, perceptions of agent reliability, trust in the agent and attitudes toward the agent. Understanding how the nuances of supplying and interpreting information that the agent will use for learning and updating its algorithms will affect the human teammate is important for agent learning after deployment, as it instructs developers how to handle future agent learning needs.

Demonstration and intervention learning are also future capabilities expected to be introduced as part of Army modernization efforts. This study offers insight as to how having the human teammate intervene and directly teach the learning agent during a mission influences team performance. As compared to when the agent inferred the information needed to alter its programming, intervening and teaching the agent resulted in increased response times, greater workload, reduced performance satisfaction, reduced perceived agent reliability, and reduced trust in the agent. As this is a future ability that could be exploited to achieve overmatch in the field, more research is required before this technology could be safely and reliably deployed without negatively impacting team performance.

7. References

- Bartneck C, Kulić D, Croft E, Zoghbi S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*. 2008;1(1):71–81. doi:10.1007/s12369-008-0001-3.
- Bhaskara A, Duong L, Brooks J, Li R, McInerney R, Skinner M, Pongracic H, Loft S. Effect of automation transparency in the management of multiple unmanned vehicles. *Applied Ergonomics*. 2021;90:103243.
- Birrell J, Meares K, Wilkinson A, Freeston M. Toward a definition of intolerance of uncertainty: a review of factor analytical studies of the intolerance of uncertainty scale. *Clinical Psychology Review*. 2011;31(7):1198–1208.
- Bitan Y, Meyer J. Self-initiated and respondent actions in a simulated control task. *Ergonomics*. 2007;50(5):763–788.
- Boyes ME, French DJ. Neuroticism, stress, and coping in the context of an anagram-solving task. *Personality and Individual Differences*. 2010;49(5):380–385.
- Breazeal C, Thomaz AL. Learning from human teachers with socially guided exploration. *IEEE International Conference on Robotics and Automation*. 2008;3539-3544. doi: 10.1109/ROBOT.2008.4543752.
- Buhr, K., Dugas, M. The intolerance of uncertainty scale: psychometric properties of the English version. *Behaviour Research and Therapy*, 2002;40(8), 931–945. doi:10.1016/s0005-7967(01)00092-4.
- Carleton RN, Norton MPJ, Asmundson GJ. Fearing the unknown: a short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders*. 2007;21(1):105–117.
- Chen JY, Barnes MJ. Supervisory control of multiple robots in dynamic tasking environments. *Ergonomics*. 2012;55(9):1043–1058.
- Chen JY, Barnes MJ. Human–agent teaming for multirobot control: a review of human factors issues. *IEEE Transactions on Human-Machine Systems*. 2014;44(1):13–29. doi: 10.1109/THMS.2013.2293535.
- Chen JY, Barnes MJ, Qu Z, Snyder MG. (2010). RoboLeader: an intelligent agent for enhancing supervisory control of multiple robots. Army Research Laboratory (US): 2010. Report No.: ARL-TR-5239. <https://doi.org/10.21236/ADA514855>.

- Chen JY, Procci K, Boyce M, Wright JL, Garcia A, Barnes MJ. Situation Awareness-Based Agent Transparency. Army Research Laboratory (US): 2014. Report No.: ARL-TR-6905.
- Chen JYC, Lakhmani SG, Stowers K, Selkowitz AR, Wright JL, Barnes MJ. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*. 2018;19(3):259–282. <https://doi.org/10.1080/1463922X.2017.1315750>.
- Chien SY, Sycara K, Liu JS, Kumru A. Relation between trust attitudes toward automation, Hofstede's cultural dimensions, and big five personality traits. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2016 Sep; Los Angeles, CA. SAGE Publications. 2016;60(1):841–845.
- Cohen MS, Parasuraman R, Freeman J. Trust in decision aids: a model and its training implications. *Cognitive Technologies*; 1999. Report No.: USAATCOM TR 97-1-4.
- Dadashi N, Stedmon AW, Pridmore TP. Semi-automated CCTV surveillance: the effects of system confidence, system accuracy and task complexity on operator vigilance, reliance and workload. *Applied Ergonomics*. 2013;44(5):730–738.
- David RA, Nielsen P. Defense science board summer study on autonomy. Defense Science Board Washington United States; 2016 June 1.
- de Visser EJ, Cohen M, Freedy A, Parasuraman R. A design methodology for trust cue calibration in cognitive agents. In: *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments: 6th International Conference, VAMR 2014, Held as Part of HCI International 2014*; 2014 June 22–27; Heraklion, Crete, Greece. Springer International Publishing; 2014. p. 251–262. *Proceedings, Part I 6 2014*.
- de Visser EJ, Monfort SS, McKendrick R, Smith MA, McKnight PE, Krueger F, Parasuraman R. Almost human: anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*. 2016;22(3):331.
- Dennett DC. True believers: the intentional strategy and why it works. In: O'Connor T, Robb D, editors. *Philosophy of mind: contemporary readings*. 2003; Routledge, London. p. 370–390.
- Derryberry D, Reed MA. Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology*. 2002;111(2):225–236. doi:10.1037/0021-843x.111.2.225.

- Deutsch M. Trust and suspicion. *Journal of Conflict Resolution*. 1958;2:265–279.
- Dongshin K, Sang MO, Rehg JM. Traversability classification for UGV navigation: A comparison of patch and superpixel representations. *IEEE/RSJ International Conference on Intelligent Robots and Systems*; 2007; San Diego, CA. p. 3166–3173. doi: 10.1109/IROS.2007.4399610.
- Donnellan MB, Oswald FL, Baird BM, Lucas RE. The mini-IPIP scales: tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment*. 2006;18:192–203.
- Dugas MJ, Freeston MH, Ladouceur R. Intolerance of uncertainty and problem orientation in worry. *Cognitive Therapy and Research*. 1997;21(6):593–606.
- Endsley MR. Situation awareness global assessment technique (SAGAT). *Proceeding of the National Aerospace and Electronics Conference (NAECON)*. IEEE; 1988. p. 789–795.
- Endsley MR. Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 1995;37(1):65-84. doi:10.1518/001872095779049499.
- Endsley MR. Level of automation: integrating humans and automated systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 1997;41(1):200–204. doi:10.1177/107118139704100146.
- Endsley MR. Direct measurement of situation awareness: validity and use of SAGAT. *Situational Awareness*. 2017;129–156. doi:10.4324/9781315087924-9.
- Evans AM, Revelle W. Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*. 2008;42(6):1585–1593.
- Goff M, Ackerman PL. Personality-intelligence relations: assessment of typical intellectual engagement. *Journal of Educational Psychology*. 1992;84(4):537.
- Goldberg JH, Kotval XP. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*. 1999;24(6):631–645.
- Greco V, Roger D. Coping with uncertainty: the construction and validation of a new measure. *Personality and Individual Differences*. 2001;31(4):519–534.
- Greco V, Roger D. Uncertainty, stress, and health. *Personality and Individual Differences*. 2003;34(6):1057–1068.
- Hamacher A, Bianchi-Berthouze N, Pipe AG, Eder K. Believing in BERT: using expressive communication to enhance trust and counteract operational error in

- physical human-robot interaction. In: 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE; 2016 Aug. p. 493–500.
- Hart SG, Staveland LE. Development of NASA-TLX (Task Load INDEX): results of empirical and theoretical research. *Advances in Psychology*. 1988;139–183. doi:10.1016/s0166-4115(08)62386-9.
- Hinds P, Roberts T, Jones H. Whose job is it anyway? A study of human robot interaction in a collaborative task. *Human-Computer Interaction*. 2004;19(1):151–181. http://dx.doi.org/10.1207/s15327051hci901&2_7.
- Ishihara S. Tests for color-blindness. Hongo Harukicho; 1917.
- Jian JY, Bissantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*. 2000;4(1):53–71. doi:10.1207/S15327566IJCE0401_04.
- Jones D, Kaber D. Situation awareness measurement and the situation awareness global assessment technique. *Handbook of Human Factors and Ergonomics Methods*. 2004;445–453. doi:10.1201/9780203489925-53.
- Kilgore R, Voshell M. Increasing the transparency of unmanned systems: applications of ecological interface design. *Proceedings of the 6th International Conference on Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality*; 2014;8526:378–389.
- Kim D, Oh SM, Rehg JM. Traversability classification for UGV navigation: a comparison of patch and superpixel representations. In: 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2007 Oct 29. IEEE; 2007. p. 3166–317.
- Lee JD, See KA. Trust in automation: designing for appropriate reliance. *Human Factors*. 2004;46(1):50–80. https://doi.org/10.1518/hfes.46.1.50_30392.
- Lin J, Matthews G, Wohleber RW, Funke GJ, Calhoun GL, Ruff HA, Chiu P. Overload and automation-dependence in a multi-UAS simulation: task demand and individual difference factors. *Journal of Experimental Psychology: Applied*. 2020;26(2):218.
- Lyons J. Being transparent about transparency: a model for human-robot interaction. *AAAI Spring Symposium*. 2013. Technical Report. 48-53.
- Lyons JB, Mahoney S, Wynne KT, Roebke MA. Viewing machines as teammates: a qualitative study. *2018 AAAI Spring Symposium Series*; 2018.

- Lyons JB, Sadler GG, Koltai K, Battiste H, Ho NT, Hoffmann LC, Smith D, Johnson W, Shively R. Shaping trust through transparent design: theoretical and experimental guidelines. In: *Advances in Human Factors in Robots and Unmanned Systems: Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems*; 2016 July 27–31; Walt Disney World, Florida. Springer International Publishing; 2017. p. 127–136.
- Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers J A. The influence of shared mental models on team process and performance. *Journal of Applied Psychology*. 2000;85(2):273–283. doi: 10.1037/0021-9010.85.2.273.
- Matthews G. Personality and information processing: a cognitive-adaptive theory. *Handbook of Personality Theory and Assessment*. 2008;1:56–79.
- Matthews G, Hancock PA, Lin J, Panganiban AR, Reinerman-Jones LE, Szalma JL, Wohleber RW. Evolution and revolution: personality research for the coming world of robots, artificial intelligence, and autonomous systems. *Personality and Individual Differences*. 2021;169:109969.
- McCrae RR, Costa PTJ. Personality trait structure as a human universal. *American Psychologist*. 1997;52(5):509–516.
- Mercado JE, Rupp MA, Chen JY, Barber D, Procci K, Barnes M. Effects of agent transparency on multi-robot management effectiveness. *Army Research Laboratory (US)*; 2015 Sep. Report No.: ARL-TR-7466.
- Mercado JE, Rupp MA, Chen JYC, Barnes MJ, Barber D, Procci K. Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 2016;58.3(2016): 401–415. doi: 10.1177/0018720815621206.
- Merritt S, Ilgen D. Not all trust is created equal: dispositional and history-based trust in human-automation interactions. *Human Factors*. 2008;50:194–210. 10.1518/001872008×288574.
- Merritt SM, Heimbaugh H, LaChapell J, Lee D. I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system. *Human factors*. 2013;55(3):520–534.
- Merritt SM, Unnerstall JL, Lee D, Huber, K. Measuring individual differences in the perfect automation schema. *Human Factors*. 2015;57(5):740–753.
- Muir BM. Trust between humans and machines, and the design of decision aids. In: Hollnagel E, Mancini G, Woods DD, editors. *Cognitive engineering in complex dynamic worlds*. Academic; 1988. p. 71–83.

- Nass C, Moon Y. Machines and mindlessness: social responses to computers. *Journal of Social Issues*. 2000;56(1):81–103.
- Nomura T, Suzuki T, Kanda T, Kato. Measurement of anxiety toward robots. In RO-MAN'07. Proceeding of the 14th IEEE international symposium on robot and human interactive communication; 2007; IEEE Press; pp. 372–377. <https://doi.org/10.1109/ROMAN.2006.314462>.
- O'Neill T, McNeese N, Barron A, Schelble B. Human–autonomy teaming: a review and analysis of the empirical literature. *Human Factors*. 2022;64(5), 904–938 doi:10.1177/0018720820960865.
- Ososky S, Sanders T, Jentsch F, Hancock P, Chen JY. Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. *Proc. SPIE 9084, Unmanned Systems Technology XVI*; 2014 June 3. p. 112–123.
- Pak R, Crumley-Branyon JJ, de Visser EJ, Rovira E. Factors that affect younger and older adults' causal attributions of robot behaviour. *Ergonomics*. 2020;63(4):421–439.
- Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. *Human Factors*. 1997;39(2):230–253.
- Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*. 2000;30(3):286–297. doi: 10.1109/3468.844354.
- Parasuraman R, Sheridan TB, Wickens CD. Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*. 2008; 2(2):140–160.
- Riley V. A general model of mixed-initiative human-machine systems. *Proceedings of the 33rd Annual Meeting of the Human Factors Society*; 1989. Denver, CO. Human Factors Society; 1989. p. 124–128.
- Robinette P, Li W, Allen R, Howard A, Wagner AR. Overtrust of robots in emergency evacuation scenarios. *IEEE International Conference on Human–Robot Interaction*; 2016. doi:10.1109/HRI.2016.7451740.
- Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: a cross-discipline view of trust. *Academy of Management Review*. 1998;23(3):393–404.

- Russell S, Norvig P. Artificial intelligence: a modern approach. Prentice Hall Press; 2009.
- Salem M, Lakatos G, Amirabdollahian F, Dautenhahn K. Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE; 2015. p. 1–8.
- Salmon PM, Stanton NA, Walker GH, Jenkins D, Ladvá D, Rafferty L, Young M. Measuring situation awareness in complex systems: comparison of measures study. *International Journal of Industrial Ergonomics*. 2009;39(3):490–500.
- Sammer G. Heart period variability and respiratory changes associated with physical and mental load: non-linear analysis. *Ergonomics*. 1998;41(5):746–755
- Sanders T, Kaplan A, Koch R, Schwartz M, Hancock PA. The relationship between trust and use choice in human-robot interaction. *Human Factors*. 2019;61(4):614–626.
- Schaefer KE, Straub ER, Chen JY, Putney J, Evans A. Communicating intent to develop shared situation awareness and engender trust in human–agent teams. *Cognitive Systems Research*. 2017;46:26-39. doi:10.1016/j.cogsys.2017.02.002.
- Selkowitz AR, Lakhmani SG, Larios CN, Chen JY. Agent transparency and the autonomous squad member. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2016 Sep; Los Angeles, CA. SAGE Publications. 2016;60(1):1319–1323.
- Seppelt BD, Lee JD. Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies*. 2007;65(3):192–205.
- Servaas MN, Riese H, Renken RJ, Marsman JBC, Lambregs J, Ormel J, Aleman A. The effect of criticism on functional brain connectivity and associations with neuroticism. *PLOS One*. 2013;8(7):e69606.
- Stanton NA, Young MS, Walker GH. (2007). The psychology of driving automation: a discussion with Professor Don Norman. *International Journal of Vehicle Design*. 2007;45(3):289–306.
- Syrdal DS, Dautenhahn K, Woods SN, Walters ML, Koay KL. Looking good? Appearance preferences and robot personality inferences at zero acquaintance.

- In: AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics. 2007;86:230–234.
- Szalma JL, Taylor GS. Individual differences in response to automation: the five factor model of personality. *Journal of Experimental Psychology: Applied*. 2011;17(2):71.
- Taylor RM. Trust and awareness in human-electronic crew teamwork. *Proceedings of the Conference on the Human-Electronic Crew: Is the Team Maturing?* 1990; Ingolstadt, Germany. p. 99–103.
- Wang H, Lewis M, Velaapudi P, Scerri P, Sycara K. How search and its subtasks scale in N robots. *Proceedings of the 4th ACM/IMEE International Conference on Human Robot Interaction*; 2009. p.141–148.
- Weld DS, Bansal G. The challenge of crafting intelligible intelligence. *Communications of the ACM*. 2019;62(6):70-79.
- Wooldridge M, Jennings NR. *Intelligent agents: theory and practice*. Knowledge Engineering Review. 1995;10(2):115–152.
- Wright JL, Chen JY, Barnes MJ, Hancock PA. Agent reasoning transparency: the influence of information level on automation induced complacency. *Army Research Laboratory (US)*: 2017. Report No.: ARL TR-8044.
- Wright JL, Chen JY, Lakhmani SG. Agent transparency and reliability in human-robot interaction: the influence on user confidence and perceived reliability. *IEEE Trans Hum Mach Syst*. 2020a;50:254–263.
- Wright JL, Chen JYC, Barnes MJ. Human-automation interaction for multiple robot control: the effect of varying automation assistance and individual differences on operator performance. *Ergonomics*. 2018;61(8):1033–1045.
- Wright JL, Chen JYC, Lakhmani SG, Selkowitz AR. Agent transparency for an autonomous squad member: depth of reasoning and reliability. *Army Research Laboratory (US)*: 2020b. Report No.: ARL-TR-8956.
- Wright JL, Chen JYC, Quinn SA, Barnes MJ. The effects of level of autonomy on human-agent teaming for multi-robot control and local security maintenance. *Army Research Laboratory (US)*; 2013. Report No.: ARL-TR-6724
- Wright JL, Lee J, Schreck JA. Human-autonomy teaming with learning capable agents: performance and workload outcomes. *International Conference on Applied Human Factors and Ergonomics*; 2021. Springer, Cham. p. 3–10.

- Wykowska A, Kajopoulous J, Obando-Leiton M, Chauhan SS, Cabibihan J, Cheng G. Humans are well tuned to detecting agents among non-agents: examining the sensitivity of human perception to behavioral characteristics of intentional systems. *International Journal of Social Robotics*. 2015;7(5):767–781. <https://doi.org/10.1007/s12369-015-0299-6>.
- Wynne KT, Lyons JB. An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*. 2018;19(3):353–374.
- Yeh YY, Wickens CD. Dissociation of performance and subjective measures of workload. *Human Factors*. 1988;30:111–120.

Appendix A. Threat Classification Factors and Training

The logic used in the simulation is scripted and the participant is thoroughly trained to understand the classifications before beginning the experimental trials. The images here are slides from the participants' classification training.

Non-Human or Human

- Occasionally, non-human objects may make it difficult for the AiTR agent to correctly classify
- Only humans can be identified as a potential threat
 - Your team's task is to identify potential human threats



Non-human



Human

Non-Human



Human



Military or Civilian

- Military denotes allied military forces within the city and are not considered a potential threat.
- Civilians are those who are part of the populace and may pose a potential threat.
- Occasionally civilians will wear an article of clothing that is camo, and the agent will have difficulty correctly classifying.



United States



United Nations



Civilian – human
in camo



Civilian

Military



Civilian



Non-Affiliated or Affiliated

- Non-affiliated people are those who are not part of a discernable group.
- Affiliated people are those who are part of a discernable group (e.g., military, red beret).
 - These civilians are a potential threat

Non-Affiliated



Affiliated



Non-Affiliated



Affiliated



Affiliated



Non-Group or Group

- A non-group is fewer than 3 people, or persons who are not interacting but merely in the proximity of others (e.g. waiting for a bus).
- Groups are three or more people
 - A group is described as people who are mutually interacting, such as talking or walking together.
- If someone affiliated with the military is part of the group, it is not considered a threat.

Non-Group



Group



3 people (group)
non-threat



3 people (group)
non-threat



3 people (group)
potential threat



Unarmed or Armed

- Unarmed is when someone either has nothing in their hands or are carrying something non-threatening such as a broom or rake.
- Armed is when someone has or is within reach of a weapon.
 - If the weapon is with military personnel or on a worksite, it is not a threat.
- Weapons include:
 - Any type of firearm, axe, and/or grenade vest

Unarmed



Armed



Unarmed



Unarmed (broom)



Armed (grenade vest)



Armed (axe)



Armed (rifle)

AiTR Reasoning Icons



Non-Human
Vs
Human



Military
Vs
Civilian



Non-Affiliated
Vs
Affiliated



Non-group
Vs
Group



Unarmed
Vs
Armed



AiTR Logic Overview

- Recall that the AiTR algorithm for classifying persons is very simple:
 - It uses image recognition to determine if what it detects is “more like” one of two options.
 - It can be imagined as a line with an option at either end. As features are recognized (or not) the needle is moved toward one end or the other.
 - The closer to the center of the line, the more evenly weighted the two options are.
 - The more evenly weighted the two options, the more likely that the agent will not be able to make a determination on that factor.**



AiTR Reasoning Icons

- When the agent makes a determination on a factor, it will highlight the associated icon
- When the AiTR cannot make a determination the icon will remain grayed out.
- The AiTR will use the information you supply during classifications to improve its ability to make determinations.
- You must also classify the potential threat and submit the reports.



Appendix B. Demographics Questionnaire

This appendix appears in its original form, without editorial change.

Demographic Questionnaire

1. General Information

- a. Age: _____ Gender: _____ Handedness: _____
- b. Do you routinely use corrective vision device(s)? (Circle one) Yes No
- i. If so, what do you use? Glasses Contact Lenses
- ii. Are you wearing them today? Yes No
- c. How many hours did you sleep last night? _____ hours

2. Military Experience

- a. Do you have prior military service? YES NO If Yes, how long _____

3. Educational Data

- a. What is your highest level of education completed or degree received? Select one.
- | | |
|--------------------------------------|-------------------------|
| _____ GED | _____ Bachelor's Degree |
| _____ High School | _____ M.S/M.A |
| _____ Some College | _____ Ph.D. |
| _____ Associates or Technical Degree | |

4. Computer Experience

- a. How often do you play computer/video games? Do not include phone/tablet games. (Circle one)
- Daily Weekly Monthly Once or twice a year Never
- b. Enter the names of the games you play most frequently:

- c. How often do you operate a radio-controlled vehicle (car, boat, plane, drone)?
- Daily Weekly Monthly Once or twice a year Never

Appendix C. Ishihara Color Test

The Ishihara Color Test,¹ 9-plate version, was administered via computer. Participants were shown images similar to that shown in Fig. C-1, and they were instructed to read the number in the display. Two incorrect answers disqualified the person from participating in the experiment.

The computer-administered version of the Ishihara was evaluated and found to be similar in efficacy to the original plates.^{2,3}

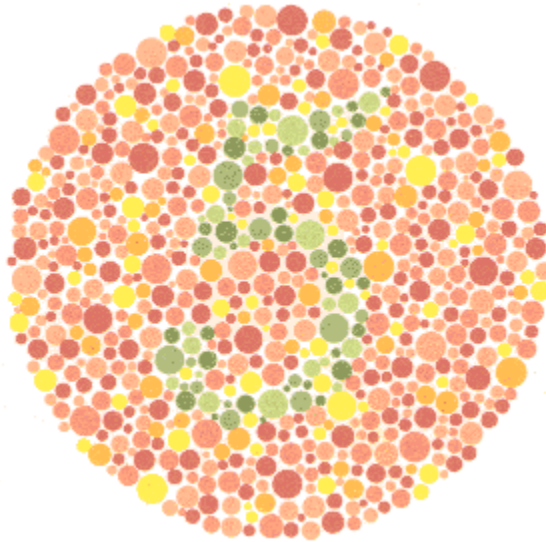


Fig. C-1 Example of Ishihara color test image

¹ Ishihara S. Tests for color-blindness. Hongo Harukicho; 1917.

² Hoffmann A, Menozz, M. Applying the Ishihara test to a PC-based screening system. Displays. 1999;20(1):39–47.

³ Pardo PJ, Pérez AL, Suero MI. Validity of TFT-LCD displays for colour vision deficiency research and diagnosis. Displays. 2004;25(4):159–163.

Appendix D. NASA-Task Load Index (TLX)

This appendix appears in its original form, without editorial change.

NASA-TLX Questionnaire

Please rate your overall impression of demands imposed on you during the exercise.

1. Mental Demand: How much mental and perceptual activity was required (e.g., thinking, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?



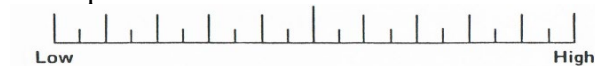
2. Physical Demand: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?



3. Temporal Demand: How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?



4. Level of Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?



5. Level of Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?



6. Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?



Appendix E. Trust in Automation

This appendix appears in its original form, without editorial change.

Please evaluate the AiTR agent you teamed with on this mission. Consider only this AiTR agent, no others, when completing the following assessment.

1 = not at all; 7 = extremely

1.	The AiTR agent is deceptive.					
1	2	3	4	5	6	7
2.	The AiTR agent behaves in an underhanded manner.					
1	2	3	4	5	6	7
3.	I am suspicious of the intent, action, or outputs of the AiTR agent.					
1	2	3	4	5	6	7
4.	I am wary of the AiTR agent.					
1	2	3	4	5	6	7
5.	The actions of the AiTR agent will have harmful or injurious outcomes.					
1	2	3	4	5	6	7
6.	I am confident in the AiTR agent.					
1	2	3	4	5	6	7
7.	The AiTR agent provides security.					
1	2	3	4	5	6	7
8.	The AiTR agent has integrity.					
1	2	3	4	5	6	7
9.	The AiTR agent is dependable.					
1	2	3	4	5	6	7
10.	The AiTR agent is reliable.					
1	2	3	4	5	6	7
11.	I can trust the AiTR agent.					
1	2	3	4	5	6	7
12.	I am familiar with the AiTR agent.					
1	2	3	4	5	6	7

Appendix F. Godspeed Questionnaire Survey

This appendix appears in its original form, without editorial change.

ROBOT ASSESSMENT SCALE

Please rate your impression of the robot on these scales:

1.	Dislike	1	2	3	4	5	Like
2.	Irresponsible	1	2	3	4	5	Responsible
3.	Inert	1	2	3	4	5	Interactive
4.	Dead	1	2	3	4	5	Alive
5.	Fake	1	2	3	4	5	Natural
6.	Anxious	1	2	3	4	5	Relaxed
7.	Apathetic	1	2	3	4	5	Responsive
8.	Agitated	1	2	3	4	5	Calm
9.	Incompetent	1	2	3	4	5	Competent
10.	Moving Rigidly	1	2	3	4	5	Moving
	Elegantly						
11.	Unkind	1	2	3	4	5	Kind
12.	Unintelligent	1	2	3	4	5	Intelligent
13.	Stagnant	1	2	3	4	5	Lively
14.	Ignorant	1	2	3	4	5	
	Knowledgeable						
15.	Machinelike	1	2	3	4	5	Humanlike
16.	Unconscious	1	2	3	4	5	Conscious
17.	Mechanical	1	2	3	4	5	Organic
18.	Artificial	1	2	3	4	5	Lifelike
19.	Awful	1	2	3	4	5	Nice
20.	Foolish	1	2	3	4	5	Sensible
21.	Quiescent	1	2	3	4	5	Surprised
22.	Unfriendly	1	2	3	4	5	Friendly
23.	Unpleasant	1	2	3	4	5	Pleasant
24.	Artificial	1	2	3	4	5	Lifelike

Appendix G. Perceived Attentional Control Survey

This appendix appears in its original form, without editorial change.

For each of the following questions, circle the response that best describes you.

It is very hard for me to concentrate on a difficult task when there are noises around.

Almost never, Sometimes,

Often, Always

When I need to concentrate and solve a problem, I have trouble focusing my attention.

Almost never, Sometimes,

Often, Always

When I am working hard on something, I still get distracted by events around me.

Almost never, Sometimes,

Often, Always

My concentration is good even if there is music in the room around me.

Almost never, Sometimes,

Often, Always

When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me. Almost

never, Sometimes, Often, Always

When I am reading or studying, I am easily distracted if there are people talking in the same room.

Almost never, Sometimes,

Often, Always

When trying to focus my attention on something, I have difficulty blocking out distracting thoughts.

Almost never, Sometimes,

Often, Always

I have a hard time concentrating when I'm excited about something.

Almost never, Sometimes,

Often, Always

When concentrating, I ignore feelings of hunger or thirst.

Almost never, Sometimes,

Often, Always

I can quickly switch from one task to another.

Almost never, Sometimes,

Often, Always

It takes me a while to get really involved in a new task.

Almost never, Sometimes,

Often, Always

It is difficult for me to coordinate my attention between the listening and writing required when taking notes during lectures.

Almost never, Sometimes, Often, Always

I can become interested in a new topic very quickly when I need to.

Almost never, Sometimes,

Often, Always

It is easy for me to read or write while I'm also talking on the phone.

Almost never, Sometimes,

Often, Always

I have trouble carrying on two conversations at once.

Almost never, Sometimes,

Often, Always

I have a hard time coming up with new ideas quickly.

Almost never, Sometimes,

Often, Always

After being interrupted or distracted, I can easily shift my attention back to what I was doing before.

Almost never, Sometimes,

Often, Always

When a distracting thought comes to mind, it is easy for me to shift my attention away from it.

Almost never, Sometimes,

Often, Always

It is easy for me to alternate between two different tasks.

Almost never, Sometimes,

Often, Always

It is hard for me to break from one way of thinking about something and look at it from another point of view.

Almost never, Sometimes,

Often, Always

Appendix H. Mini-International Personality Item Pool (IPIP)

This appendix appears in its original form, without editorial change.

Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement whether it is 1. Very Inaccurate, 2. Moderately Inaccurate, 3. Neither Accurate Nor Inaccurate, 4. Moderately Accurate, or 5. Very Accurate as a description of you.

	Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate
1. Am the life of the party.	☐	☐	☐	☐	☐
2. Sympathize with others' feelings	☐	☐	☐	☐	☐
3. Get chores done right away.	☐	☐	☐	☐	☐
4. Have frequent mood swings.	☐	☐	☐	☐	☐
5. Have a vivid imagination.	☐	☐	☐	☐	☐
6. Don't talk a lot.	☐	☐	☐	☐	☐
7. Am not interested in other people's problems.	☐	☐	☐	☐	☐
8. Often forget to put things back in their proper place.	☐	☐	☐	☐	☐
9. Am relaxed most of the time.	☐	☐	☐	☐	☐
10. Am not interested in abstract ideas.	☐	☐	☐	☐	☐
11. Talk to a lot of different people at parties.	☐	☐	☐	☐	☐
12. Feel others' emotions.	☐	☐	☐	☐	☐
13. Like order.	☐	☐	☐	☐	☐
14. Get upset easily.	☐	☐	☐	☐	☐
15. Have difficulty understanding abstract ideas.	☐	☐	☐	☐	☐
16. Keep in the background.	☐	☐	☐	☐	☐
17. Am not really interested in others.	☐	☐	☐	☐	☐
18. Make a mess of things.	☐	☐	☐	☐	☐
19. Seldom feel blue.	☐	☐	☐	☐	☐
20. Do not have a good imagination.	☐	☐	☐	☐	☐

Appendix I. Intolerance of Uncertainty

This appendix appears in its original form, without editorial change.

Describe yourself as you generally are now, not as you wish to be in the future or may have been in the past. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement select the response that best describes you: 1. Almost Never, 2. Sometimes, 3. Often, or 4. Always.

	Almost Never	Sometimes	Often	Always
Unforeseen events upset me greatly.	☐	☐	☐	☐
It frustrates me not having all the information I need.	☐	☐	☐	☐
One should always look ahead so as to avoid surprises.	☐	☐	☐	☐
A small, unforeseen event can spoil everything, even with the best of planning.	☐	☐	☐	☐
I always want to know what the future has in store for me.	☐	☐	☐	☐
I can't stand being taken by surprise.	☐	☐	☐	☐
I should be able to organize everything in advance.	☐	☐	☐	☐
Uncertainty keeps me from living a full life.	☐	☐	☐	☐
When it's time to act, uncertainty paralyses me.	☐	☐	☐	☐
When I am uncertain I can't function very well.	☐	☐	☐	☐
The smallest doubt can stop me from acting.	☐	☐	☐	☐
I must get away from all uncertain situations.	☐	☐	☐	☐

Appendix J. Situation Awareness Global Assessment Technique (SAGAT) Probes

This appendix appears in its original form, without editorial change.

Four times each trial the simulation was paused, the screen blackened, and the following four questions (examples shown below) were presented. Questions 1, 2, and 5 assessed the participants' SA1 (perception), questions 3 & 4 assessed their SA2 (comprehension). The questions were designed to sample between the primary task (threat classification) duties and the secondary task (maintain awareness of vehicle route and environment), as a comprehensive understanding of the state of the entire operational environment is required when performing this function in the field. As such, these queries have face validity with the needs of a Soldier performing the same tasking.

1. Which potential threat did you most recently encounter?
 - a. Armed civilian walking
 - b. Armed faction member standing
 - c. Armed civilian at bus stop
 - d. Faction member walking
2. What did the route planner rank the current route's safety?
 - a. Good (>70%)
 - b. Fair (30 – 70%)
 - c. Poor (<30%)
3. How well does the AiTR classify potential threats (do not include undetermined persons)?
 - a. Very reliable (no errors)
 - b. Mostly reliable (few errors)
 - c. Unreliable (many errors)

The final question varied between these two options:

4. Has your vehicle altered its route, and if so what was the most recent reason why? (SA2)
 - d. No, Still on original route
 - e. Yes, Encountered closed road
 - f. Yes, potential IED in roadway
 - g. Yes, Communications down in area
5. In what area of the city is your vehicle currently? (SA1)_
 - h. Central, near the park
 - i. South, near the soccer field
 - j. East, near the marketplace
 - k. North, above the park

Appendix K. University of Central Florida (UCF) COVID Information Sheet and Screening Questionnaire

This appendix appears in its original form, without editorial change.

**UCF****Important Information about COVID-19 and Research Participation**

At the University of Central Florida our primary responsibility related to research is to protect the safety of our research participants.

COVID-19 refers to the Coronavirus that is being spread across people in our communities. We need to provide you with important information about COVID-19, and to tell you about ways your study participation might change because of COVID-19 related risk.

If you are considering joining a study at this time or are currently enrolled in a study, it is important that you consider the following information to determine if study participation is right for you at this time.

How is COVID-19 spread? COVID-19 is a respiratory virus spread by respiratory droplets, mainly from person-to-person. This can happen between people who are in close contact with one another (less than 6 feet). It is also possible that a person can get COVID-19 by touching a surface or object (such as a doorknob or counter surface) that has the virus on it, then touching their mouth, nose or eyes.

Can COVID-19 be prevented? Current ways to minimize the risk of exposure to COVID-19 include "social distancing" which is a practice to decrease the potential for direct exposure to others who may have been exposed to COVID-19, for example by avoiding large gatherings or refraining from shaking hands with others. It is important to understand that since study participation may include increased travel outside of your home and increased exposure to others within a research site it may increase your exposure to COVID-19. Currently, there is no vaccination to prevent COVID-19 infection.

What are the risks of COVID-19? For most people, the new coronavirus causes only mild or moderate symptoms, such as fever and cough. For some, especially older adults and people with existing health problems, it can cause more severe illness, including pneumonia. While we are still learning about this virus, the information we have right now suggests that about 3 out of every 100 people who are infected might die from the virus.

Who is most at risk?

- Persons over 65 years of age
- Persons with chronic lung disease or moderate to severe asthma
- Persons with diabetes
- Persons with serious heart conditions
- Persons with severe obesity (body mass index of 40 or higher)
- Persons with liver disease
- Persons with chronic kidney disease undergoing dialysis
- Immunocompromised individuals (cancer treatment, bone marrow/organ recipients, those with chronic immunodeficiencies, poorly controlled HIV or AIDS, prolonged use of corticosteroids and other immune weakening medications)

How could your participation in this research change as a result of COVID-19? There are several ways we try to minimize your risk. If possible, we limit the number of times you have to come

University of Central Florida v 6/12/2020

to a research site. We will ask every research participant if they have the symptoms of COVID-19 or have been in close contact with anyone who has or had COVID-19 using the following questions:

Health Screening Questions:

1. IN THE PAST 24 HOURS, have you had any of the following symptoms? YES NO

- Fever
- Cough
- Sore throat
- Shortness of Breath
- Loss of smell or taste
- Chills or repeated shaking with chills
- Muscle fatigue
- Headache
- Nausea, vomiting, or diarrhea
- Congestion or runny nose

If "YES", DO NOT participate in this study.

2. Have you TRAVELED INTERNATIONALLY in the past 14 days? YES NO

If "YES", DO NOT participate in this study.

3. Have you TRAVELED DOMESTICALLY (U.S) to areas of high COVID-19 prevalence (e.g., New York City, south Florida, etc.) in the past 14 days? YES NO

If "YES", DO NOT participate in this study.

4. Have you had CLOSE PERSONAL CONTACT with anyone who has been diagnosed with COVID-19 in the past 14 days (per criteria below)? YES NO

- a. Within 6 feet for prolonged period of time
- b. In direct contact with infectious secretions (been coughed/sneezed upon, etc.)

If "YES", DO NOT participate in this study.

During your research visits, we try to reduce the time you are exposed to other people as much as possible. You will be asked to wear a cloth face covering upon arrival to the research location and during most, if not all, of the research procedures. Please let the study team know if you need a face covering provided to you or if you can bring your own. Your study team will review the Informed Consent Form with you and discuss if there are any study procedures during which you will be asked to briefly remove the covering or use a different type of face covering.

If you are suspected to be positive for COVID-19, there may be last minute changes to how research procedures are performed [such as a change from an in-person visit to a telephone call] or cancellations of research tests or procedures to ensure your safety. It is even possible that your research procedures will be put on hold or stopped because of COVID-19.

University of Central Florida v 6/12/2020

The information related to risks of COVID-19 changes every day. The leaders at UCF are monitoring these risks and deciding how these risks should change our research. If you have questions about COVID-19 and your participation in research, please talk to your study team.

You will find a hotline to use to report any concerns you have with this study related to COVID-19 [here](#).

Please contact the principal investigator of this study if you develop symptoms of COVID-19 or test positive within two weeks of your study participation.

Adapted with permission from the Johns Hopkins University.

Appendix L. Data Tables

Task Performance Measures

Table L-1 Descriptive statistics for performance measures

	T1			T2			T3			T4		
	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
Overall Correct IDs (%)	91.34	4.80	1.07	93.75	3.85	0.86	94.56	2.84	0.64	93.49	3.75	0.84
Correct Classifications (%)	82.20	7.33	0.16	83.30	9.07	0.20	85.10	5.19	0.12	84.20	7.65	0.17
Correct Acceptances (%)	93.20	4.59	0.10	95.90	3.73	0.08	96.30	2.77	0.08	95.60	3.62	0.08
Overall Response Time	10.89	2.54	0.57	9.66	1.89	0.42	8.29	1.39	0.31	14.03	4.95	1.11
Response Time - Correct IDs	10.80	2.47	0.55	9.54	1.86	0.42	8.21	1.35	0.30	13.89	5.02	1.12
Response Time - Incorrect IDs	12.35	4.90	1.09	12.29	4.67	1.04	9.57	3.01	0.67	19.04	5.85	1.31

Table L-2 Paired t-test results for performance measures

	T1 - T2			T2 - T3			T3 - T4		
	<i>t</i> (19)	<i>p</i>	<i>d_s</i>	<i>t</i> (19)	<i>p</i>	<i>d_s</i>	<i>t</i> (19)	<i>p</i>	<i>d_s</i>
Overall Correct IDs (%)	-2.44	0.025	-0.55	-0.96	0.348	-0.24	1.55	0.137	0.32
Correct Classifications (%)	-0.54	0.594	0.13	-0.55	0.352	0.25	0.61	0.547	0.13
Correct Acceptances (%)	-2.66	0.016	0.66	-0.39	0.699	0.11	0.92	0.367	0.22
Overall Response Time	2.28	0.034	0.55	4.35	< .001	0.83	-5.95	< .001	-1.58
Response Time - Correct IDs	2.46	0.024	0.57	4.33	< .001	0.82	-5.85	< .001	-1.55
Response Time - Incorrect IDs	0.04	0.968	0.01	2.24	0.037	0.69	-6.84	< .001	-2.03

Table L-3 Descriptive statistics by task load for each transparency condition

		T1			T2			T3			T4		
		<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
Correct IDs (%)	High TL	89.10	6.45	1.44	92.00	5.33	1.19	92.65	3.70	0.83	91.96	4.92	1.10
	Low TL	94.49	3.99	0.89	95.69	3.44	0.77	96.22	3.71	0.83	95.99	3.21	0.72
Response Time	High TL	10.56	2.25	0.50	9.53	1.92	0.43	8.17	1.24	0.28	19.01	11.58	2.59
	Low TL	10.61	2.86	0.64	9.59	2.74	0.61	8.20	1.60	0.36	12.08	3.33	0.74

Table L-4 T-test results high vs. low task load outcomes within each task load condition

	Correct IDs (%)			Overall Response Time		
	t(19)	<i>p</i>	<i>d_s</i>	t(19)	<i>p</i>	<i>d_s</i>
T1	-4.56	< .001	-1.01	-0.09	0.931	-0.02
T2	-3.42	0.003	-0.82	-0.09	0.928	-0.02
T3	-3.67	0.002	-0.96	-0.08	0.939	-0.02
T4	-4.32	< .001	-0.97	3.42	0.003	0.81

NASA-TLX Workload Measures

Table L-5 Descriptive statistics for NASA-TLX results

	T1			T2			T3			T4		
	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
Overall	50.29	16.52	3.69	44.63	17.57	3.93	41.33	16.69	3.73	53.46	16.91	3.78
MD	75.00	16.38	3.66	59.25	25.56	5.72	55.00	24.71	5.53	67.25	24.03	5.37
PD	34.75	27.41	6.13	32.75	27.12	6.06	33.75	26.40	5.90	37.00	28.76	6.43
TD	61.00	26.49	5.92	56.00	24.09	5.39	51.00	27.61	6.17	70.50	18.98	4.24
Effort	65.75	15.15	3.39	54.00	24.31	5.44	51.75	21.42	4.79	66.75	20.98	4.69
Frustration	38.50	29.11	6.51	32.25	25.10	5.61	29.75	23.59	5.28	38.25	26.72	5.97
Perf. Satis.	31.00	17.96	4.02	32.00	18.74	4.19	25.25	14.64	3.27	41.00	23.82	5.33

Table L-6 Paired T-test results for NASA-TLX scores

	T1 - T2			T2 - T3			T3 - T4		
	t(19)	<i>p</i>	<i>d_s</i>	t(19)	<i>p</i>	<i>d_s</i>	t(19)	<i>p</i>	<i>d_s</i>
Overall	2.02	0.06	0.33	1.92	0.07	0.19	-4.54	0.00	-0.72
MD	4.54	0.00	0.73	1.34	0.20	0.17	-3.39	0.00	-0.50
PD	0.44	0.66	0.07	-0.57	0.58	-0.04	-1.09	0.29	-0.12
TD	1.23	0.23	0.20	1.37	0.19	0.19	-3.96	0.00	-0.82
Effort	2.66	0.02	0.58	0.55	0.59	0.10	-5.43	0.00	-0.71
Frustration	1.29	0.21	0.23	1.56	0.14	0.10	-1.81	0.09	-0.34
Perf. Satis.	-0.29	0.77	-0.05	2.93	0.01	0.40	-4.47	0.00	-0.80

Situation Awareness and Perceived Reliability Measures

Table L-7 Situation awareness (SA) and perceived reliability scores descriptive statistics. LTL – low task load. HTL – high task load.

		T1			T2			T3			T4		
		<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
SA1	Overall	0.70	0.10	0.02	0.67	0.16	0.04	0.67	0.16	0.04	0.65	0.16	0.04
	LTL	0.62	0.19	0.04	0.69	0.20	0.04	0.68	0.21	0.05	0.64	0.19	0.04
	HTL	0.75	0.18	0.04	0.64	0.23	0.05	0.66	0.21	0.05	0.68	0.16	0.04
SA2	Overall	0.33	0.19	0.04	0.43	0.24	0.05	0.48	0.24	0.05	0.42	0.25	0.06
	LTL	0.41	0.23	0.05	0.51	0.22	0.05	0.56	0.21	0.05	0.50	0.23	0.05
	HTL	0.09	0.15	0.03	0.41	0.27	0.06	0.30	0.38	0.08	0.25	0.38	0.09
Perceived Reliability	Overall	1.15	0.38	0.08	1.21	0.28	0.06	1.31	0.38	0.09	1.11	0.51	0.11
	LTL	1.18	0.47	0.10	1.43	0.27	0.06	1.35	0.43	0.10	1.13	0.53	0.12
	HTL	1.13	0.36	0.08	1.40	0.27	0.06	1.28	0.38	0.08	1.10	0.55	0.12

Table L-8 Paired t-test results for situation awareness and perceived reliability results

	T1 - T2			T2 - T3			T3 - T4		
	t(19)	p	d _s	t(19)	p	d _s	t(19)	p	d _s
SA1	0.88	0.389	0.24	0.17	0.582	0.03	0.56	0.582	0.16
SA2	1.22	0.237	0.42	1.14	0.267	0.21	1.28	0.217	0.24
Perceived. Reliability	0.53	0.601	0.19	1.29	0.214	0.30	2.56	0.019	0.45

Table L-9 Paired t-test results for SA and perceived reliability scores by task load (low LTL vs. high HTL)

	T1			T2			T3			T4		
	t(19)	p	d _s	t(19)	p	d _s	t(19)	p	d _s	t(19)	p	d _s
SA1	2.02	0.057	0.70	0.81	0.431	0.22	0.31	0.760	0.09	0.65	0.527	0.18
SA2	5.38	0.000	1.66	2.03	0.057	0.40	3.80	0.001	0.78	3.68	0.002	0.73
Perceived. Reliability	0.62	0.541	0.12	1.00	0.330	0.09	1.14	0.267	0.18	0.30	0.772	0.05

Trust and Godspeed Measures

Table L-10 Descriptive statistics for trust and godspeed questionnaires

	T1			T2			T3			T4		
	M	SD	SE	M	SD	SE	M	SD	SE	M	SD	SE
Trust	55.60	12.68	2.83	58.65	11.74	2.62	60.45	11.51	2.57	53.75	10.21	2.28
Anthropomorphism	2.78	0.62	0.14	2.92	0.80	0.18	2.83	0.93	0.21	2.54	0.69	0.15
Animacy	3.05	0.80	0.18	3.17	0.83	0.18	3.18	0.95	0.21	2.89	0.84	0.19
Likeability	3.52	0.70	0.16	3.56	0.66	0.15	3.49	0.86	0.19	3.16	0.67	0.15
Perceived Intelligence	4.08	0.50	0.11	4.12	0.47	0.10	3.96	0.86	0.19	3.56	0.65	0.15
Perceived Safety	3.37	0.61	0.14	3.28	0.42	0.09	3.38	0.65	0.15	3.21	0.43	0.10

Table L-11 Paired t-test results for Trust and Godspeed questionnaires between transparency conditions

	T1 - T2			T2 - T3			T3 - T4		
	t(19)	p	d _s	t(19)	p	d _s	t(19)	p	d _s
Trust	-1.34	0.197	-0.25	-1.07	0.300	-0.15	2.90	0.009	0.62
Anthropomorphism	-1.06	0.303	-0.20	0.62	0.545	0.10	1.59	0.129	0.35
Animacy	-0.99	0.332	-0.14	-0.18	0.861	-0.02	1.49	0.152	0.33
Likeability	-0.30	0.766	-0.06	0.58	0.569	0.09	1.61	0.125	0.43
Perceived Intelligence	-0.42	0.677	-0.08	0.88	0.389	0.23	1.68	0.109	0.52
Perceived Safety	0.68	0.507	0.16	-1.00	0.329	-0.18	1.35	0.192	0.31

Individual Difference Factors

Perceived Attentional Control

Table L-12 Performance statistics by high/low perceived attentional control

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Correct IDs (mean %)	T1	HI	10	91.36	4.77	1.51	0.02	0.987	0.01
		LO	10	91.33	5.09	1.61			
	T2	HI	10	94.82	3.70	1.17	1.26	0.225	0.56
		LO	10	92.69	3.88	1.23			
	T3	HI	10	94.61	3.30	1.04	0.07	0.942	0.03
		LO	10	94.51	2.49	0.79			
	T4	HI	10	94.55	3.37	1.06	1.29	0.214	0.58
		LO	10	92.43	3.99	1.26			
Response Time	T1	HI	10	10.95	2.79	0.88	0.10	0.921	0.04
		LO	10	10.83	2.42	0.77			
	T2	HI	10	9.78	1.86	0.59	0.28	0.780	0.13
		LO	10	9.54	2.01	0.63			
	T3	HI	10	8.48	1.49	0.47	0.59	0.561	0.26
		LO	10	8.10	1.32	0.42			
	T4	HI	10	12.96	5.10	1.61	-0.97	0.345	-0.42
		LO	10	15.11	4.82	1.52			
Response Time - Correct	T1	HI	10	10.90	2.63	0.83	0.18	0.862	0.08
		LO	10	10.70	2.43	0.77			
	T2	HI	10	9.69	1.83	0.58	0.34	0.737	0.15
		LO	10	9.40	1.97	0.62			
	T3	HI	10	8.40	1.36	0.43	0.62	0.542	0.28
		LO	10	8.02	1.38	0.44			
	T4	HI	10	12.79	5.16	1.63	-0.98	0.341	-0.44
		LO	10	14.99	4.89	1.55			
Response Time - Incorrect	T1	HI	10	11.36	5.13	1.62	-0.90	0.381	-0.40
		LO	10	13.33	4.70	1.49			
	T2	HI	10	12.11	4.94	1.56	-0.17	0.870	-0.07
		LO	10	12.46	4.64	1.47			
	T3	HI	10	9.74	3.79	1.20	0.23	0.817	0.10
		LO	10	9.41	2.17	0.69			
	T4	HI	10	18.06	5.54	1.75	-0.74	0.469	-0.33
		LO	10	20.02	6.28	1.99			

Table L-13 Workload factor scores by high/low perceived attentional control

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Overall Workload	T1	LO	10	49.40	18.83	5.95	-0.48	0.639	-0.21
		HI	10	53.00	14.61	4.62			
	T2	LO	10	44.20	19.19	6.07	-0.10	0.922	-0.04
		HI	10	45.00	16.86	5.33			
	T3	LO	10	38.90	14.68	4.64	-0.67	0.511	-0.30
		HI	10	44.00	19.08	6.04			
	T4	LO	10	49.00	15.61	4.94	-1.19	0.250	-0.53
		HI	10	57.90	17.80	5.63			
Mental Demand	T1	LO	10	71.00	17.76	5.62	-1.10	0.287	-0.49
		HI	10	79.00	14.68	4.64			
	T2	LO	10	56.50	29.25	9.25	-0.47	0.643	-0.21
		HI	10	62.00	22.51	7.12			
	T3	LO	10	48.50	24.73	7.82	-1.19	0.250	-0.53
		HI	10	61.50	24.16	7.64			
	T4	LO	10	56.50	25.17	7.96	-2.19	0.042	-0.98
		HI	10	78.00	18.14	5.73			
Physical Demand	T1	LO	10	28.00	23.83	7.54	-1.11	0.282	-0.50
		HI	10	41.50	30.28	9.58			
	T2	LO	10	25.00	25.71	8.13	-1.30	0.210	-0.58
		HI	10	40.50	27.53	8.71			
	T3	LO	10	24.00	26.54	8.39	-1.74	0.099	-0.78
		HI	10	43.50	23.58	7.46			
	T4	LO	10	27.00	27.10	8.57	-1.62	0.123	-0.72
		HI	10	47.00	28.11	8.89			
Temporal Demand	T1	LO	10	68.00	24.18	7.64	1.20	0.248	0.53
		HI	10	54.00	28.07	8.88			
	T2	LO	10	58.50	26.04	8.23	0.45	0.655	0.20
		HI	10	53.50	23.10	7.30			
	T3	LO	10	50.00	19.15	6.06	-0.16	0.876	-0.07
		HI	10	52.00	35.21	11.14			
	T4	LO	10	66.50	17.17	5.43	-0.94	0.360	-0.42
		HI	10	74.50	20.74	6.56			
Effort	T1	LO	10	62.00	17.03	5.39	-1.11	0.280	-0.50
		HI	10	69.50	12.79	4.04			
	T2	LO	10	50.50	27.13	8.58	-0.63	0.534	-0.28
		HI	10	57.50	22.02	6.96			
	T3	LO	10	49.50	20.47	6.47	-0.46	0.651	-0.21
		HI	10	54.00	23.19	7.33			
	T4	LO	10	60.50	22.17	7.01	-1.36	0.190	-0.61
		HI	10	73.00	18.74	5.93			

Table L-13 Workload factor scores by high/low perceived attentional control (continued)

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Frustration	T1	LO	10	43.50	31.18	9.86	0.76	0.457	0.34
		HI	10	33.50	27.59	8.73			
	T2	LO	10	37.50	26.69	8.44	0.93	0.364	0.42
		HI	10	27.00	23.59	7.46			
	T3	LO	10	32.50	22.52	7.12	0.51	0.615	0.23
		HI	10	27.00	25.52	8.07			
	T4	LO	10	43.00	23.83	7.54	0.79	0.441	0.35
		HI	10	33.50	29.82	9.43			
Performance Satisfaction	T1	LO	10	27.50	16.87	5.34	-0.87	0.398	-0.39
		HI	10	34.50	19.21	6.08			
	T2	LO	10	35.50	20.06	6.34	0.83	0.418	0.37
		HI	10	28.50	17.65	5.58			
	T3	LO	10	26.50	15.47	4.89	0.37	0.713	0.17
		HI	10	24.00	14.49	4.58			
	T4	LO	10	40.00	18.56	5.87	-1.19	0.250	-0.53
		HI	10	42.00	29.17	9.23			

Table L-14 Situation awareness scores by high/low perceived attentional control

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Situation Awareness Level 1	T1	LO	10	0.70	0.10	0.03	0.11	0.914	0.05
		HI	10	0.70	0.10	0.03			
	T2	LO	10	0.61	0.19	0.06	1.60	0.127	0.72
		HI	10	0.72	0.11	0.04			
	T3	LO	10	0.62	0.15	0.05	1.44	0.166	0.65
		HI	10	0.72	0.15	0.05			
	T4	LO	10	0.57	0.15	0.05	2.32	0.032	1.04
		HI	10	0.72	0.14	0.04			
Situation Awareness Level 2	T1	LO	10	0.40	0.22	0.07	1.66	0.113	0.74
		HI	10	0.27	0.12	0.04			
	T2	LO	10	0.37	0.23	0.07	1.07	0.299	0.48
		HI	10	0.48	0.25	0.08			
	T3	LO	10	0.42	0.16	0.05	1.11	0.284	0.49
		HI	10	0.53	0.29	0.09			
	T4	LO	10	0.32	0.23	0.07	1.90	0.073	0.85
		HI	10	0.52	0.24	0.08			
Perceived Reliability	T1	LO	10	1.25	0.47	0.15	1.21	0.243	0.54
		HI	10	1.05	0.23	0.07			
	T2	LO	10	1.15	0.24	0.08	0.98	0.339	0.44
		HI	10	1.28	0.32	0.10			
	T3	LO	10	1.28	0.34	0.11	0.43	0.671	0.15
		HI	10	1.35	0.43	0.14			
	T4	LO	10	0.93	0.53	0.17	1.73	0.101	0.77
		HI	10	1.30	0.44	0.14			

Table L-15 Jian TIA scale scores by high/low perceived attentional control

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Trust in the Agent	T1	LO	10	53.60	10.50	3.32	-0.70	0.495	-0.31
		HI	10	57.60	14.83	4.69			
	T2	LO	10	55.50	12.89	4.07	-1.22	0.240	-0.54
		HI	10	61.80	10.13	3.20			
	T3	LO	10	57.90	10.40	3.29	-0.99	0.335	-0.44
		HI	10	63.00	12.53	3.96			
	T4	LO	10	48.40	8.71	2.75	-2.71	0.014	-1.21
		HI	10	59.10	8.97	2.84			

Table L-16 Godspeed survey scores by high/low perceived attentional control

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Anthropomorphism	T1	LO	10	2.78	0.63	0.20	0.00	1.000	0.00
		HI	10	2.78	0.64	0.20			
	T2	LO	10	2.86	0.62	0.20	-0.33	0.746	-0.15
		HI	10	2.98	0.97	0.31			
	T3	LO	10	2.60	0.54	0.17	-1.11	0.283	-0.50
		HI	10	3.06	1.20	0.38			
	T4	LO	10	2.48	0.42	0.13	-0.38	0.708	-0.17
		HI	10	2.60	0.90	0.29			
Animacy	T1	LO	10	2.99	0.79	0.25	-0.36	0.720	-0.16
		HI	10	3.12	0.85	0.27			
	T2	LO	10	3.05	0.64	0.20	-0.62	0.541	-0.28
		HI	10	3.28	1.00	0.32			
	T3	LO	10	3.13	0.72	0.23	-0.23	0.821	-0.10
		HI	10	3.23	1.17	0.37			
	T4	LO	10	2.72	0.54	0.17	-0.93	0.365	-0.42
		HI	10	3.07	1.07	0.34			
Likeability	T1	LO	10	3.44	0.44	0.14	-0.50	0.624	-0.22
		HI	10	3.60	0.91	0.29			
	T2	LO	10	3.20	0.44	0.14	-2.86	0.010	-1.28
		HI	10	3.92	0.66	0.21			
	T3	LO	10	3.14	0.57	0.18	-1.95	0.067	-0.87
		HI	10	3.84	0.98	0.31			
	T4	LO	10	2.86	0.21	0.07	-2.19	0.042	-0.98
		HI	10	3.46	0.84	0.27			
Perceived Intelligence	T1	LO	10	4.00	0.45	0.14	-0.71	0.489	-0.32
		HI	10	4.16	0.56	0.18			
	T2	LO	10	3.98	0.43	0.13	-1.38	0.185	-0.62
		HI	10	4.26	0.48	0.15			
	T3	LO	10	4.06	0.61	0.19	0.51	0.616	0.23
		HI	10	3.86	1.08	0.34			
	T4	LO	10	3.24	0.55	0.17	-2.48	0.023	-1.11
		HI	10	3.88	0.61	0.19			
Perceived Safety	T1	LO	10	3.30	0.46	0.14	-0.48	0.636	-0.22
		HI	10	3.43	0.75	0.24			
	T2	LO	10	3.00	0.31	0.10	-4.01	0.001	-1.79
		HI	10	3.57	0.32	0.10			
	T3	LO	10	3.03	0.40	0.13	-2.81	0.012	-1.26
		HI	10	3.74	0.68	0.22			
	T4	LO	10	3.03	0.29	0.09	-1.96	0.066	-0.87
		HI	10	3.39	0.49	0.16			

Intolerance of Uncertainty

Table L-17 Performance measures by high/low prospective anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Correct IDs (mean %)	T1	HI	10	91.76	5.81	1.84	0.38	0.712	0.17
		LO	10	90.93	3.81	1.20			
	T2	HI	10	92.58	4.69	1.48	-1.40	0.179	-0.62
		LO	10	94.93	2.49	0.79			
	T3	HI	10	93.95	2.84	0.90	-0.96	0.349	-0.43
		LO	10	95.17	2.86	0.90			
	T4	HI	10	91.92	3.96	1.25	-2.01	0.060	-0.90
		LO	10	95.05	2.92	0.92			
Response Time	T1	HI	10	10.76	2.72	0.86	-0.22	0.827	-0.10
		LO	10	11.02	2.49	0.79			
	T2	HI	10	9.11	1.85	0.58	-1.32	0.204	-0.59
		LO	10	10.20	1.86	0.59			
	T3	HI	10	8.08	1.32	0.42	-0.66	0.518	-0.30
		LO	10	8.50	1.49	0.47			
	T4	HI	10	15.53	4.81	1.52	1.38	0.184	0.62
		LO	10	12.54	4.87	1.54			
Response Time - Correct	T1	HI	10	10.67	2.68	0.85	-0.22	0.828	-0.10
		LO	10	10.92	2.38	0.75			
	T2	HI	10	9.00	1.83	0.58	-1.34	0.196	-0.60
		LO	10	10.09	1.80	0.57			
	T3	HI	10	8.02	1.40	0.44	-0.63	0.538	-0.28
		LO	10	8.40	1.34	0.42			
	T4	HI	10	15.36	4.95	1.57	1.33	0.200	0.60
		LO	10	12.43	4.89	1.55			
Response Time - Incorrect	T1	HI	10	13.00	5.35	1.69	0.59	0.562	0.26
		LO	10	11.69	4.58	1.45			
	T2	HI	10	11.25	4.37	1.38	-0.99	0.336	-0.44
		LO	10	13.32	4.95	1.57			
	T3	HI	10	9.44	2.38	0.75	-0.19	0.850	-0.09
		LO	10	9.71	3.66	1.16			
	T4	HI	10	21.10	5.39	1.70	1.64	0.118	0.73
		LO	10	16.98	5.82	1.84			

Table L-18 Performance measures by high/low inhibitory anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Correct IDs (mean %)	T1	HI	11	91.03	4.88	1.47	-0.32	0.753	-0.14
		LO	9	91.73	4.97	1.66			
	T2	HI	11	92.36	4.26	1.29	-1.91	0.072	-0.86
		LO	9	95.46	2.56	0.85			
	T3	HI	11	94.40	2.96	0.89	-0.28	0.786	-0.12
		LO	9	94.76	2.86	0.95			
	T4	HI	11	92.71	4.32	1.30	-1.03	0.315	-0.46
		LO	9	94.45	2.88	0.96			
Response Time	T1	HI	11	10.78	2.79	0.84	-0.21	0.834	-0.10
		LO	9	11.03	2.37	0.79			
	T2	HI	11	9.16	2.02	0.61	-1.33	0.200	-0.60
		LO	9	10.27	1.62	0.54			
	T3	HI	11	7.85	1.09	0.33	-1.64	0.118	-0.74
		LO	9	8.83	1.58	0.53			
	T4	HI	11	14.58	4.81	1.45	0.54	0.597	0.24
		LO	9	13.36	5.33	1.78			
Response Time - Correct	T1	HI	11	10.57	2.65	0.80	-0.44	0.667	-0.20
		LO	9	11.07	2.34	0.78			
	T2	HI	11	9.03	1.97	0.59	-1.42	0.174	-0.64
		LO	9	10.18	1.59	0.53			
	T3	HI	11	7.81	1.16	0.35	-1.54	0.140	-0.69
		LO	9	8.71	1.46	0.49			
	T4	HI	11	14.45	4.87	1.47	0.54	0.597	0.24
		LO	9	13.21	5.41	1.80			
Response Time - Incorrect	T1	HI	11	13.62	5.58	1.68	1.31	0.207	0.59
		LO	9	10.79	3.62	1.21			
	T2	HI	11	11.96	4.65	1.40	-0.34	0.741	-0.15
		LO	9	12.68	4.94	1.65			
	T3	HI	11	8.70	2.33	0.70	-1.48	0.155	-0.67
		LO	9	10.64	3.52	1.17			
	T4	HI	11	19.74	6.07	1.83	0.58	0.570	0.26
		LO	9	18.19	5.82	1.94			

Table L-19 Workload factor scores by high/low prospective anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Overall Workload	T1	LO	10	54.80	13.18	4.17	0.97	0.343	0.44
		HI	10	47.60	19.30	6.10			
	T2	LO	10	45.90	17.00	5.38	0.32	0.751	0.14
		HI	10	43.30	18.97	6.00			
	T3	LO	10	41.10	19.43	6.14	-0.09	0.929	-0.04
		HI	10	41.80	14.71	4.65			
	T4	LO	10	54.60	17.27	5.46	0.30	0.770	0.13
		HI	10	52.30	17.41	5.51			
Mental Demand	T1	LO	10	80.50	11.89	3.76	1.56	0.137	0.70
		HI	10	69.50	18.92	5.98			
	T2	LO	10	66.00	20.39	6.45	1.19	0.248	0.53
		HI	10	52.50	29.37	9.29			
	T3	LO	10	59.50	26.82	8.48	0.81	0.430	0.36
		HI	10	50.50	22.91	7.24			
	T4	LO	10	72.50	23.83	7.54	0.98	0.342	0.44
		HI	10	62.00	24.29	7.68			
Physical Demand	T1	LO	10	37.00	28.89	9.13	0.36	0.724	0.16
		HI	10	32.50	27.21	8.60			
	T2	LO	10	29.00	25.47	8.06	-0.61	0.551	-0.27
		HI	10	36.50	29.54	9.34			
	T3	LO	10	30.50	23.15	7.32	-0.54	0.596	-0.24
		HI	10	37.00	30.20	9.55			
	T4	LO	10	38.00	26.27	8.31	0.15	0.881	0.07
		HI	10	36.00	32.47	10.27			
Temporal Demand	T1	LO	10	59.50	28.33	8.96	-0.25	0.808	-0.11
		HI	10	62.50	25.95	8.21			
	T2	LO	10	55.50	24.99	7.90	-0.09	0.929	-0.04
		HI	10	56.50	24.50	7.75			
	T3	LO	10	50.00	35.90	11.35	-0.16	0.876	-0.07
		HI	10	52.00	17.83	5.64			
	T4	LO	10	72.50	16.71	5.28	0.46	0.650	0.21
		HI	10	68.50	21.74	6.87			
Effort	T1	LO	10	70.50	11.65	3.69	1.44	0.167	0.64
		HI	10	61.00	17.29	5.47			
	T2	LO	10	60.00	21.08	6.67	1.11	0.281	0.50
		HI	10	48.00	26.89	8.50			
	T3	LO	10	49.50	24.77	7.83	-0.46	0.651	-0.21
		HI	10	54.00	18.53	5.86			
	T4	LO	10	68.50	22.24	7.03	0.36	0.720	0.16
		HI	10	65.00	20.68	6.54			

Table L-19 Workload factor scores by high/low prospective anxiety (continued)

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Frustration	T1	LO	10	34.50	29.76	9.41	-0.60	0.553	-0.27
		HI	10	42.50	29.46	9.32			
	T2	LO	10	27.00	27.41	8.67	-0.93	0.364	-0.42
		HI	10	37.50	22.76	7.20			
	T3	LO	10	26.00	27.67	8.75	-0.70	0.492	-0.31
		HI	10	33.50	19.44	6.15			
	T4	LO	10	29.50	28.43	8.99	-1.51	0.148	-0.68
		HI	10	47.00	23.00	7.27			
Performance Satisfaction	T1	LO	10	37.00	21.76	6.88	1.55	0.139	0.69
		HI	10	25.00	11.30	3.57			
	T2	LO	10	35.50	22.66	7.17	0.83	0.418	0.37
		HI	10	28.50	14.15	4.48			
	T3	LO	10	27.50	16.54	5.23	0.68	0.507	0.30
		HI	10	23.00	12.95	4.10			
	T4	LO	10	47.00	31.11	9.84	1.14	0.271	0.51
		HI	10	35.00	12.25	3.87			

Table L-20 Workload factor scores by high/low inhibitory anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Overall Workload	T1	LO	9	52.67	14.69	4.90	0.35	0.730	0.16
		HI	11	50.00	18.48	5.57			
	T2	LO	9	47.11	14.84	4.95	0.57	0.578	0.25
		HI	11	42.55	20.03	6.04			
	T3	LO	9	41.33	13.43	4.48	-0.03	0.978	-0.01
		HI	11	41.55	19.76	5.96			
	T4	LO	9	52.33	14.09	4.70	-0.26	0.798	-0.12
		HI	11	54.36	19.57	5.90			
Mental Demand	T1	LO	9	81.11	14.74	4.91	1.57	0.135	0.70
		HI	11	70.00	16.58	5.00			
	T2	LO	9	65.56	20.07	6.69	1.00	0.332	0.45
		HI	11	54.09	29.22	8.81			
	T3	LO	9	57.22	24.38	8.13	0.36	0.726	0.16
		HI	11	53.18	26.01	7.84			
	T4	LO	9	65.00	23.05	7.68	-0.37	0.716	-0.17
		HI	11	69.09	25.77	7.77			
Physical Demand	T1	LO	9	26.67	20.92	6.97	-1.21	0.243	-0.54
		HI	11	41.36	31.15	9.39			
	T2	LO	9	29.44	29.31	9.77	-0.48	0.635	-0.22
		HI	11	35.45	26.31	7.93			
	T3	LO	9	28.89	28.48	9.49	-0.74	0.471	-0.33
		HI	11	37.73	25.24	7.61			
	T4	LO	9	36.11	32.09	10.70	-0.12	0.904	-0.05
		HI	11	37.73	27.33	8.24			
Temporal Demand	T1	LO	9	75.00	12.99	4.33	2.39	0.028	1.07
		HI	11	49.55	29.62	8.93			
	T2	LO	9	67.78	15.83	5.28	2.16	0.045	0.97
		HI	11	46.36	25.99	7.84			
	T3	LO	9	58.89	18.33	6.11	1.17	0.258	0.52
		HI	11	44.55	32.82	9.90			
	T4	LO	9	74.44	14.67	4.89	0.83	0.415	0.37
		HI	11	67.27	22.06	6.65			
Effort	T1	LO	9	65.56	15.50	5.17	-0.05	0.960	-0.02
		HI	11	65.91	15.62	4.71			
	T2	LO	9	56.67	24.87	8.29	0.43	0.669	0.20
		HI	11	51.82	24.83	7.49			
	T3	LO	9	49.44	18.95	6.32	-0.43	0.675	-0.19
		HI	11	53.64	23.99	7.23			
	T4	LO	9	64.44	18.78	6.26	-0.43	0.669	-0.20
		HI	11	68.64	23.35	7.04			

Table L-20 Workload factor scores by high/low inhibitory anxiety (continued)

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Frustration	T1	LO	9	37.78	29.27	9.76	-0.10	0.923	-0.04
		HI	11	39.09	30.40	9.17			
	T2	LO	9	31.67	21.94	7.31	-0.09	0.928	-0.04
		HI	11	32.73	28.49	8.59			
	T3	LO	9	28.33	19.53	6.51	-0.24	0.815	-0.11
		HI	11	30.91	27.37	8.25			
	T4	LO	9	39.44	23.91	7.97	0.18	0.862	0.08
		HI	11	37.27	29.95	9.03			
Performance Satisfaction	T1	LO	9	24.44	17.58	5.86	-1.53	0.144	-0.69
		HI	11	36.36	17.19	5.18			
	T2	LO	9	30.00	21.36	7.12	-0.42	0.678	-0.19
		HI	11	33.64	17.19	5.18			
	T3	LO	9	22.22	14.39	4.80	-0.83	0.418	-0.37
		HI	11	27.73	15.06	4.54			
	T4	LO	9	34.44	20.38	6.79	-1.12	0.277	-0.50
		HI	11	46.36	25.99	7.84			

Table L-21 Situation awareness scores by high/low prospective anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Situation Awareness Level 1	T1	LO	10	0.72	0.09	0.03	1.01	0.324	0.45
		HI	10	0.68	0.11	0.03			
	T2	LO	10	0.67	0.17	0.05	0.14	0.893	0.06
		HI	10	0.66	0.16	0.05			
	T3	LO	10	0.67	0.19	0.06	0.00	1.000	0.00
		HI	10	0.67	0.13	0.04			
	T4	LO	10	0.70	0.12	0.04	1.59	0.129	0.71
		HI	10	0.59	0.18	0.06			
Situation Awareness Level 2	T1	LO	10	0.30	0.11	0.03	0.79	0.441	0.34
		HI	10	0.37	0.25	0.08			
	T2	LO	10	0.45	0.22	0.07	0.45	0.660	0.20
		HI	10	0.40	0.27	0.09			
	T3	LO	10	0.48	0.28	0.09	0.15	0.880	0.07
		HI	10	0.47	0.20	0.06			
	T4	LO	10	0.47	0.20	0.06	0.89	0.387	0.40
		HI	10	0.37	0.29	0.09			
Perceived Reliability	T1	LO	10	1.10	0.13	0.04	0.59	0.565	0.26
		HI	10	1.20	0.52	0.17			
	T2	LO	10	1.28	0.32	0.10	0.98	0.339	0.44
		HI	10	1.15	0.24	0.08			
	T3	LO	10	1.33	0.47	0.15	0.14	0.887	0.06
		HI	10	1.30	0.28	0.09			
	T4	LO	10	1.25	0.44	0.14	0.22	0.237	0.55
		HI	10	0.98	0.56	0.18			

Table L-22 Situation awareness scores by high/low inhibitory anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Situation Awareness Level 1	T1	LO	9	0.71	0.09	0.03	0.54	0.593	0.24
		HI	11	0.69	0.11	0.03			
	T2	LO	9	0.73	0.11	0.04	1.83	0.083	0.82
		HI	11	0.61	0.18	0.05			
	T3	LO	9	0.71	0.15	0.05	1.05	0.309	0.47
		HI	11	0.64	0.16	0.05			
	T4	LO	9	0.69	0.15	0.05	1.11	0.280	0.50
		HI	11	0.61	0.17	0.05			
Situation Awareness Level 2	T1	LO	9	0.33	0.17	0.06	0.00	1.000	0.00
		HI	11	0.33	0.21	0.06			
	T2	LO	9	0.48	0.27	0.09	0.93	0.364	0.42
		HI	11	0.38	0.22	0.07			
	T3	LO	9	0.44	0.25	0.08	-0.51	0.616	0.23
		HI	11	0.50	0.24	0.07			
	T4	LO	9	0.44	0.30	0.10	0.44	0.666	0.20
		HI	11	0.39	0.21	0.06			
Perceived Reliability	T1	LO	9	1.11	0.28	0.09	-0.41	0.687	0.18
		HI	11	1.18	0.45	0.14			
	T2	LO	9	1.31	0.33	0.11	1.35	0.193	0.61
		HI	11	1.14	0.23	0.07			
	T3	LO	9	1.22	0.38	0.13	-0.96	0.349	0.43
		HI	11	1.39	0.38	0.11			
	T4	LO	9	1.14	0.56	0.19	0.20	0.841	0.09
		HI	11	1.09	0.49	0.15			

Table L-23 Jian TIA scale scores by high/low prospective anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Trust in the Agent	T1	LO	10	55.70	14.75	4.66	0.03	0.973	0.02
		HI	10	55.50	11.03	3.49			
	T2	LO	10	61.60	11.40	3.61	1.13	0.272	0.51
		HI	10	55.70	11.89	3.76			
	T3	LO	10	63.30	14.22	4.50	1.11	0.280	0.50
		HI	10	57.60	7.71	2.44			
	T4	LO	10	58.50	8.59	2.72	2.31	0.033	1.03
		HI	10	49.00	9.80	3.10			

Table L-24 Jian TIA scale scores by high/low inhibitory anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Trust in the Agent	T1	LO	9	61.22	12.33	4.11	1.92	0.071	0.86
		HI	11	51.00	11.49	3.46			
	T2	LO	9	61.44	10.84	3.61	0.96	0.349	0.43
		HI	11	56.36	12.44	3.75			
	T3	LO	9	62.56	12.89	4.30	0.73	0.474	0.33
		HI	11	58.73	10.56	3.18			
	T4	LO	9	59.00	10.63	3.54	2.30	0.033	1.04
		HI	11	49.45	7.90	2.38			

Table L-25 Godspeed survey scores by high/low prospective anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Anthropomorphism	T1	LO	10	2.78	0.72	0.23	0.00	1.000	0.00
		HI	10	2.78	0.54	0.17			
	T2	LO	10	2.98	0.97	0.31	0.33	0.746	0.15
		HI	10	2.86	0.62	0.20			
	T3	LO	10	3.04	1.22	0.39	1.01	0.328	0.45
		HI	10	2.62	0.51	0.16			
	T4	LO	10	2.62	0.88	0.28	0.51	0.617	0.23
		HI	10	2.46	0.46	0.15			
Animacy	T1	LO	10	3.04	0.96	0.30	-0.09	0.929	-0.04
		HI	10	3.07	0.65	0.21			
	T2	LO	10	3.15	1.14	0.36	-0.08	0.934	-0.04
		HI	10	3.18	0.36	0.12			
	T3	LO	10	3.17	1.29	0.41	-0.07	0.942	-0.03
		HI	10	3.20	0.49	0.15			
	T4	LO	10	3.02	1.11	0.35	0.66	0.519	0.29
		HI	10	2.77	0.46	0.15			
Likeability	T1	LO	10	3.64	0.90	0.29	0.75	0.460	0.34
		HI	10	3.40	0.44	0.14			
	T2	LO	10	3.94	0.68	0.22	3.10	0.006	1.39
		HI	10	3.18	0.37	0.12			
	T3	LO	10	3.86	1.01	0.32	2.09	0.051	0.93
		HI	10	3.12	0.48	0.15			
	T4	LO	10	3.44	0.85	0.27	2.01	0.059	0.90
		HI	10	2.88	0.21	0.07			
Perceived Intelligence	T1	LO	10	4.14	0.56	0.18	0.53	0.605	0.24
		HI	10	4.02	0.46	0.14			
	T2	LO	10	4.30	0.47	0.15	1.83	0.083	0.82
		HI	10	3.94	0.40	0.13			
	T3	LO	10	3.82	1.11	0.35	-0.72	0.481	-0.32
		HI	10	4.10	0.54	0.17			
	T4	LO	10	3.74	0.74	0.23	1.26	0.225	0.56
		HI	10	3.38	0.52	0.16			
Perceived Safety	T1	LO	10	3.33	0.78	0.25	-0.24	0.816	-0.11
		HI	10	3.40	0.41	0.13			
	T2	LO	10	3.43	0.45	0.14	1.66	0.114	0.74
		HI	10	3.13	0.36	0.11			
	T3	LO	10	3.70	0.69	0.22	2.44	0.025	1.09
		HI	10	3.07	0.44	0.14			
	T4	LO	10	3.39	0.49	0.16	1.96	0.066	0.87
		HI	10	3.03	0.29	0.09			

Table L-26 Godspeed survey scores by high/low inhibitory anxiety

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Anthropomorphism	T1	LO	9	2.84	0.65	0.22	0.41	0.685	0.19
		HI	11	2.73	0.62	0.19			
	T2	LO	9	3.07	0.93	0.31	0.74	0.471	0.33
		HI	11	2.80	0.69	0.21			
	T3	LO	9	3.04	1.14	0.38	0.93	0.367	0.42
		HI	11	2.65	0.73	0.22			
	T4	LO	9	2.67	0.93	0.31	0.73	0.473	0.33
		HI	11	2.44	0.43	0.13			
Animacy	T1	LO	9	3.08	0.94	0.31	0.12	0.907	0.05
		HI	11	3.03	0.71	0.21			
	T2	LO	9	3.26	1.02	0.34	0.45	0.661	0.20
		HI	11	3.09	0.68	0.20			
	T3	LO	9	3.22	1.17	0.39	0.16	0.876	0.07
		HI	11	3.15	0.78	0.24			
	T4	LO	9	3.11	1.12	0.37	1.06	0.303	0.48
		HI	11	2.71	0.51	0.15			
Likeability	T1	LO	9	3.76	0.82	0.27	1.39	0.182	0.62
		HI	11	3.33	0.56	0.17			
	T2	LO	9	3.91	0.70	0.23	2.41	0.027	1.08
		HI	11	3.27	0.48	0.15			
	T3	LO	9	3.78	0.99	0.33	1.39	0.182	0.62
		HI	11	3.25	0.70	0.21			
	T4	LO	9	3.44	0.87	0.29	1.82	0.086	0.82
		HI	11	2.93	0.34	0.10			
Perceived Intelligence	T1	LO	9	4.29	0.44	0.15	1.78	0.091	0.80
		HI	11	3.91	0.50	0.15			
	T2	LO	9	4.29	0.49	0.16	1.52	0.146	0.68
		HI	11	3.98	0.41	0.12			
	T3	LO	9	3.91	1.13	0.38	-0.22	0.825	-0.10
		HI	11	4.00	0.61	0.18			
	T4	LO	9	4.00	0.58	0.19	3.42	0.003	1.54
		HI	11	3.20	0.46	0.14			
Perceived Safety	T1	LO	9	3.56	0.73	0.24	1.27	0.220	0.57
		HI	11	3.21	0.48	0.14			
	T2	LO	9	3.56	0.33	0.11	3.15	0.006	1.42
		HI	11	3.06	0.36	0.11			
	T3	LO	9	3.71	0.59	0.20	2.18	0.042	0.98
		HI	11	3.12	0.60	0.18			
	T4	LO	9	3.47	0.50	0.17	2.81	0.012	1.26
		HI	11	3.00	0.21	0.06			

Personality Factors

Table L-27 Performance statistics by high/low extroversion scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Correct IDs (mean %)	T1	HI	11	89.85	4.17	1.26	-1.60	0.127	-0.72
		LO	9	93.17	5.13	1.71			
	T2	HI	11	92.80	3.46	1.04	-1.25	0.228	-0.56
		LO	9	94.92	4.18	1.39			
	T3	HI	11	93.96	3.50	1.05	-1.05	0.308	-0.47
		LO	9	95.30	1.68	0.56			
	T4	HI	11	93.72	3.30	1.00	0.30	0.768	0.13
		LO	9	93.20	4.43	1.48			
Response Time	T1	HI	11	10.88	2.76	0.83	-0.02	0.984	-0.01
		LO	9	10.90	2.42	0.81			
	T2	HI	11	9.81	1.93	0.58	0.39	0.698	0.18
		LO	9	9.47	1.94	0.65			
	T3	HI	11	8.83	1.55	0.47	2.11	0.049	0.95
		LO	9	7.62	0.81	0.27			
	T4	HI	11	15.72	5.02	1.51	1.78	0.092	0.80
		LO	9	11.97	4.25	1.42			
Response Time - Correct	T1	HI	11	10.85	2.65	0.80	0.10	0.918	0.05
		LO	9	10.73	2.38	0.79			
	T2	HI	11	9.68	1.91	0.58	0.36	0.724	0.16
		LO	9	9.38	1.89	0.63			
	T3	HI	11	8.77	1.44	0.43	2.25	0.037	1.01
		LO	9	7.53	0.88	0.29			
	T4	HI	11	15.60	5.09	1.53	1.78	0.092	0.80
		LO	9	11.80	4.30	1.43			
Response Time - Incorrect	T1	HI	11	11.54	5.33	1.61	-0.80	0.431	-0.36
		LO	9	13.33	4.41	1.47			
	T2	HI	11	12.46	4.76	1.43	0.18	0.862	0.08
		LO	9	12.08	4.84	1.61			
	T3	HI	11	9.81	3.73	1.12	0.38	0.705	0.17
		LO	9	9.28	1.99	0.66			
	T4	HI	11	21.22	5.75	1.73	1.98	0.063	0.89
		LO	9	16.38	5.05	1.68			

Table L-28 Performance statistics by high/low agreeableness scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Correct IDs (mean %)	T1	HI	9	90.82	3.84	1.28	-0.44	0.669	-0.20
		LO	11	91.78	5.62	1.69			
	T2	HI	9	95.18	3.34	1.11	1.55	0.138	0.70
		LO	11	92.59	3.99	1.20			
	T3	HI	9	95.21	2.39	0.80	0.92	0.372	0.41
		LO	11	94.03	3.18	0.96			
	T4	HI	9	94.65	4.33	1.44	1.27	0.220	0.57
		LO	11	92.54	3.10	0.93			
Response Time	T1	HI	9	10.70	2.67	0.89	-0.30	0.769	-0.13
		LO	11	11.05	2.55	0.77			
	T2	HI	9	9.74	1.87	0.62	0.17	0.865	0.08
		LO	11	9.59	1.99	0.60			
	T3	HI	9	7.89	1.20	0.40	-1.17	0.256	-0.53
		LO	11	8.61	1.49	0.45			
	T4	HI	9	11.36	2.98	0.99	-2.45	0.025	-1.10
		LO	11	16.22	5.28	1.59			
Response Time - Correct	T1	HI	9	10.48	2.49	0.83	-0.51	0.617	-0.23
		LO	11	11.06	2.54	0.77			
	T2	HI	9	9.64	1.82	0.61	0.21	0.833	0.10
		LO	11	9.46	1.97	0.59			
	T3	HI	9	7.81	1.26	0.42	-1.24	0.231	-0.56
		LO	11	8.55	1.38	0.42			
	T4	HI	9	11.26	2.98	0.99	-2.36	0.029	-1.06
		LO	11	16.05	5.42	1.64			
Response Time - Incorrect	T1	HI	9	12.82	5.72	1.91	0.38	0.705	0.17
		LO	11	11.96	4.36	1.32			
	T2	HI	9	12.27	4.82	1.61	-0.01	0.990	-0.01
		LO	11	12.30	4.78	1.44			
	T3	HI	9	9.07	2.20	0.73	-0.67	0.509	-0.30
		LO	11	9.99	3.59	1.08			
	T4	HI	9	15.92	4.19	1.40	-2.42	0.026	-1.09
		LO	11	21.60	5.93	1.79			

Table L-29 Performance statistics by high/low conscientiousness scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Correct IDs (mean %)	T1	HI	12	91.48	4.35	1.26	0.15	0.881	0.07
		LO	8	91.14	5.72	2.02			
	T2	HI	12	94.20	3.68	1.06	0.62	0.541	0.28
		LO	8	93.09	4.26	1.51			
	T3	HI	12	94.65	3.02	0.87	0.17	0.865	0.08
		LO	8	94.42	2.75	0.97			
	T4	HI	12	94.13	3.95	1.14	0.93	0.365	0.42
		LO	8	92.53	3.46	1.22			
Response Time	T1	HI	12	11.02	2.62	0.76	0.26	0.795	0.12
		LO	8	10.70	2.58	0.91			
	T2	HI	12	9.62	1.82	0.53	-0.11	0.916	-0.05
		LO	8	9.72	2.11	0.75			
	T3	HI	12	8.43	1.35	0.39	0.54	0.595	0.25
		LO	8	8.08	1.50	0.53			
	T4	HI	12	13.89	4.79	1.38	-0.15	0.881	-0.07
		LO	8	14.24	5.52	1.95			
Response Time - Correct	T1	HI	12	10.88	2.49	0.72	0.17	0.867	0.08
		LO	8	10.68	2.60	0.92			
	T2	HI	12	9.50	1.79	0.52	-0.12	0.903	-0.06
		LO	8	9.61	2.08	0.73			
	T3	HI	12	8.36	1.25	0.36	0.58	0.569	0.26
		LO	8	7.99	1.55	0.55			
	T4	HI	12	13.74	4.87	1.41	-0.16	0.871	-0.08
		LO	8	14.12	5.57	1.97			
Response Time - Incorrect	T1	HI	12	12.42	5.33	1.54	0.08	0.937	0.04
		LO	8	12.24	4.53	1.60			
	T2	HI	12	11.83	4.73	1.37	-0.53	0.605	-0.24
		LO	8	12.97	4.80	1.70			
	T3	HI	12	9.62	3.50	1.01	0.08	0.937	0.04
		LO	8	9.51	2.32	0.82			
	T4	HI	12	19.71	5.27	1.52	0.61	0.548	0.28
		LO	8	18.04	6.89	2.44			

Table L-30 Performance statistics by high/low neuroticism scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Correct IDs (mean %)	T1	HI	10	91.78	5.23	1.65	0.39	0.698	0.18
		LO	10	90.91	4.58	1.45			
	T2	HI	10	93.85	3.86	1.22	0.11	0.917	0.05
		LO	10	93.66	4.05	1.28			
	T3	HI	10	94.74	2.58	0.81	0.27	0.790	0.12
		LO	10	94.39	3.22	1.02			
	T4	HI	10	92.70	4.28	1.35	-0.94	0.358	-0.42
		LO	10	94.28	3.16	1.00			
Response Time	T1	HI	10	10.70	2.47	0.78	-0.33	0.742	-0.15
		LO	10	11.09	2.73	0.86			
	T2	HI	10	9.67	1.93	0.61	0.04	0.970	0.02
		LO	10	9.64	1.95	0.62			
	T3	HI	10	7.83	1.17	0.37	-1.53	0.144	-0.68
		LO	10	8.75	1.49	0.47			
	T4	HI	10	13.92	5.15	1.63	-0.10	0.922	-0.04
		LO	10	14.14	5.02	1.59			
Response Time - Correct	T1	HI	10	10.59	2.46	0.78	-0.37	0.714	-0.17
		LO	10	11.01	2.59	0.82			
	T2	HI	10	9.49	1.90	0.60	-0.12	0.903	-0.06
		LO	10	9.60	1.91	0.60			
	T3	HI	10	7.74	1.21	0.38	-1.65	0.117	-0.74
		LO	10	8.69	1.37	0.43			
	T4	HI	10	13.79	5.22	1.65	-0.09	0.930	-0.04
		LO	10	13.99	5.09	1.61			
Response Time - Incorrect	T1	HI	10	12.82	5.04	1.59	0.42	0.677	0.19
		LO	10	11.87	4.97	1.57			
	T2	HI	10	13.81	4.50	1.42	1.51	0.149	0.67
		LO	10	10.76	4.54	1.44			
	T3	HI	10	9.64	2.90	0.92	0.10	0.923	0.04
		LO	10	9.51	3.27	1.03			
	T4	HI	10	18.96	7.12	2.25	-0.06	0.953	-0.03
		LO	10	19.12	4.66	1.47			

Table L-31 Performance statistics by high/low openness scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Correct IDs (mean %)	T1	HI	13	90.46	4.24	1.18	-1.12	0.276	-0.53
		LO	7	92.98	5.68	2.15			
	T2	HI	13	93.35	4.18	1.16	-0.63	0.535	-0.30
		LO	7	94.51	3.32	1.25			
	T3	HI	13	94.84	2.65	0.74	0.59	0.561	0.28
		LO	7	94.04	3.33	1.26			
	T4	HI	13	93.93	4.28	1.19	0.71	0.486	0.33
		LO	7	92.66	2.60	0.98			
Response Time	T1	HI	13	10.79	2.71	0.75	-0.24	0.809	-0.11
		LO	7	11.09	2.39	0.90			
	T2	HI	13	9.29	1.86	0.52	-1.22	0.239	-0.57
		LO	7	10.35	1.87	0.71			
	T3	HI	13	7.96	0.94	0.26	-1.48	0.156	-0.69
		LO	7	8.89	1.91	0.72			
	T4	HI	13	13.20	4.87	1.35	-1.03	0.318	-0.48
		LO	7	15.58	5.10	1.93			
Response Time - Correct	T1	HI	13	10.64	2.59	0.72	-0.38	0.707	-0.18
		LO	7	11.09	2.39	0.90			
	T2	HI	13	9.17	1.81	0.50	-1.23	0.235	-0.58
		LO	7	10.23	1.87	0.71			
	T3	HI	13	7.94	0.94	0.26	-1.24	0.232	-0.58
		LO	7	8.71	1.88	0.71			
	T4	HI	13	13.08	4.88	1.35	-0.99	0.336	-0.46
		LO	7	15.40	5.30	2.00			
Response Time - Incorrect	T1	HI	13	12.05	5.23	1.45	-0.36	0.723	-0.17
		LO	7	12.90	4.55	1.72			
	T2	HI	13	11.91	4.86	1.35	-0.48	0.638	-0.22
		LO	7	12.98	4.57	1.73			
	T3	HI	13	8.60	2.50	0.69	-2.16	0.045	-1.01
		LO	7	11.39	3.20	1.21			
	T4	HI	13	17.64	5.63	1.56	-1.51	0.148	-0.71
		LO	7	21.65	5.73	2.17			

Table L-32 Workload (NASA-TLX) statistics by high/low extroversion scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Overall Workload	T1	LO	9	52.67	14.69	4.90	0.35	0.730	0.16
		HI	11	50.00	18.48	5.57			
	T2	LO	9	47.11	14.84	4.95	0.57	0.578	0.25
		HI	11	42.55	20.03	6.04			
	T3	LO	9	41.33	13.43	4.48	-0.03	0.978	-0.01
		HI	11	41.55	19.76	5.96			
	T4	LO	9	52.33	14.09	4.70	-0.26	0.798	-0.12
		HI	11	54.36	19.57	5.90			
Mental Demand	T1	LO	9	81.11	14.74	4.91	1.57	0.135	0.70
		HI	11	70.00	16.58	5.00			
	T2	LO	9	65.56	20.07	6.69	1.00	0.332	0.45
		HI	11	54.09	29.22	8.81			
	T3	LO	9	57.22	24.38	8.13	0.36	0.726	0.16
		HI	11	53.18	26.01	7.84			
	T4	LO	9	65.00	23.05	7.68	-0.37	0.716	-0.17
		HI	11	69.09	25.77	7.77			
Physical Demand	T1	LO	9	26.67	20.92	6.97	-1.21	0.243	-0.54
		HI	11	41.36	31.15	9.39			
	T2	LO	9	29.44	29.31	9.77	-0.48	0.635	-0.22
		HI	11	35.45	26.31	7.93			
	T3	LO	9	28.89	28.48	9.49	-0.74	0.471	-0.33
		HI	11	37.73	25.24	7.61			
	T4	LO	9	36.11	32.09	10.70	-0.12	0.904	-0.05
		HI	11	37.73	27.33	8.24			
Temporal Demand	T1	LO	9	75.00	12.99	4.33	2.39	0.028	1.07
		HI	11	49.55	29.62	8.93			
	T2	LO	9	67.78	15.83	5.28	2.16	0.045	0.97
		HI	11	46.36	25.99	7.84			
	T3	LO	9	58.89	18.33	6.11	1.17	0.258	0.52
		HI	11	44.55	32.82	9.90			
	T4	LO	9	74.44	14.67	4.89	0.83	0.415	0.37
		HI	11	67.27	22.06	6.65			

Table L-32 Workload (NASA-TLX) statistics by high/low extroversion scores (continued)

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Effort	T1	LO	9	65.56	15.50	5.17	-0.05	0.960	-0.02
		HI	11	65.91	15.62	4.71			
	T2	LO	9	56.67	24.87	8.29	0.43	0.669	0.20
		HI	11	51.82	24.83	7.49			
	T3	LO	9	49.44	18.95	6.32	-0.43	0.675	-0.19
		HI	11	53.64	23.99	7.23			
	T4	LO	9	64.44	18.78	6.26	-0.43	0.669	-0.20
		HI	11	68.64	23.35	7.04			
Frustration	T1	LO	9	37.78	29.27	9.76	-0.10	0.923	-0.04
		HI	11	39.09	30.40	9.17			
	T2	LO	9	31.67	21.94	7.31	-0.09	0.928	-0.04
		HI	11	32.73	28.49	8.59			
	T3	LO	9	28.33	19.53	6.51	-0.24	0.815	-0.11
		HI	11	30.91	27.37	8.25			
	T4	LO	9	39.44	23.91	7.97	0.18	0.862	0.08
		HI	11	37.27	29.95	9.03			
Performance Satisfaction	T1	LO	9	24.44	17.58	5.86	-1.53	0.144	-0.69
		HI	11	36.36	17.19	5.18			
	T2	LO	9	30.00	21.36	7.12	-0.42	0.678	-0.19
		HI	11	33.64	17.19	5.18			
	T3	LO	9	22.22	14.39	4.80	-0.83	0.418	-0.37
		HI	11	27.73	15.06	4.54			
	T4	LO	9	34.44	20.38	6.79	-1.12	0.277	-0.50
		HI	11	46.36	25.99	7.84			

Table L-33 Workload (NASA-TLX) statistics by high/low agreeableness scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Overall Workload	T1	LO	11	54.00	15.59	4.70	0.83	0.416	0.37
		HI	9	47.78	17.86	5.95			
	T2	LO	11	48.45	16.64	5.02	1.09	0.290	0.49
		HI	9	39.89	18.52	6.17			
	T3	LO	11	44.00	16.48	4.97	0.74	0.467	0.33
		HI	9	38.33	17.58	5.86			
	T4	LO	11	57.45	19.25	5.80	1.18	0.252	0.53
		HI	9	48.56	12.95	4.32			
Mental Demand	T1	LO	11	76.36	17.33	5.23	0.40	0.692	0.18
		HI	9	73.33	16.01	5.34			
	T2	LO	11	60.91	23.96	7.22	0.31	0.758	0.14
		HI	9	57.22	28.73	9.58			
	T3	LO	11	55.00	24.29	7.32	0.00	1.000	0.00
		HI	9	55.00	26.69	8.90			
	T4	LO	11	66.82	28.22	8.51	-0.09	0.932	-0.04
		HI	9	67.78	19.38	6.46			
Physical Demand	T1	LO	11	45.91	23.11	6.97	2.21	0.040	0.99
		HI	9	21.11	27.13	9.04			
	T2	LO	11	51.36	21.69	6.54	5.26	0.000	2.37
		HI	9	10.00	10.00	3.33			
	T3	LO	11	50.00	23.45	7.07	4.14	0.001	1.86
		HI	9	13.89	12.69	4.23			
	T4	LO	11	52.27	28.32	8.54	3.20	0.005	1.44
		HI	9	18.33	15.81	5.27			
Temporal Demand	T1	LO	11	60.91	28.44	8.58	-0.02	0.987	-0.01
		HI	9	61.11	25.59	8.53			
	T2	LO	11	54.09	24.06	7.26	-0.38	0.706	-0.17
		HI	9	58.33	25.37	8.46			
	T3	LO	11	45.00	29.66	8.94	-1.08	0.295	-0.49
		HI	9	58.33	24.49	8.16			
	T4	LO	11	67.27	22.29	6.72	-0.83	0.415	-0.37
		HI	9	74.44	14.24	4.75			

Table L-33 Workload (NASA-TLX) statistics by high/low agreeableness scores (continued)

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Effort	T1	LO	11	69.09	15.46	4.66	1.10	0.288	0.49
		HI	9	61.67	14.58	4.86			
	T2	LO	11	56.82	23.05	6.95	0.56	0.581	0.25
		HI	9	50.56	26.74	8.91			
	T3	LO	11	54.09	18.95	5.71	0.53	0.603	0.24
		HI	9	48.89	24.97	8.32			
	T4	LO	11	65.91	24.48	7.38	-0.19	0.849	-0.09
		HI	9	67.78	17.16	5.72			
Frustration	T1	LO	11	41.36	30.09	9.07	0.48	0.640	0.21
		HI	9	35.00	29.26	9.75			
	T2	LO	11	31.82	26.39	7.96	-0.08	0.935	-0.04
		HI	9	32.78	25.01	8.34			
	T3	LO	11	30.45	21.96	6.62	0.14	0.887	0.06
		HI	9	28.89	26.78	8.93			
	T4	LO	11	45.91	28.18	8.50	1.46	0.162	0.66
		HI	9	28.89	22.88	7.63			
Performance Satisfaction	T1	LO	11	28.64	10.98	3.31	-0.64	0.530	-0.29
		HI	9	33.89	24.47	8.16			
	T2	LO	11	34.55	15.72	4.74	0.66	0.517	0.30
		HI	9	28.89	22.47	7.49			
	T3	LO	11	26.82	13.47	4.06	0.52	0.610	0.23
		HI	9	23.33	16.58	5.53			
	T4	LO	11	46.82	21.60	6.51	1.22	0.237	0.55
		HI	9	33.89	25.71	8.57			

Table L-34 Workload (NASA-TLX) statistics by high/low conscientiousness scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Overall Workload	T1	LO	8	52.63	21.27	7.52	0.31	0.762	0.14
		HI	12	50.25	13.42	3.88			
	T2	LO	8	48.75	18.84	6.66	0.86	0.403	0.39
		HI	12	41.83	16.95	4.89			
	T3	LO	8	43.38	15.17	5.36	0.41	0.687	0.19
		HI	12	40.17	18.31	5.29			
	T4	LO	8	51.88	14.41	5.09	-0.33	0.744	-0.15
		HI	12	54.50	18.96	5.47			
Mental Demand	T1	LO	8	76.25	19.78	6.99	0.27	0.789	0.12
		HI	12	74.17	14.59	4.21			
	T2	LO	8	60.63	29.21	10.33	0.19	0.850	0.09
		HI	12	58.33	24.15	6.97			
	T3	LO	8	54.38	26.65	9.42	-0.09	0.929	-0.04
		HI	12	55.42	24.54	7.08			
	T4	LO	8	59.38	22.59	7.99	-1.21	0.242	-0.55
		HI	12	72.50	24.45	7.06			
Physical Demand	T1	LO	8	30.00	23.60	8.35	-0.62	0.541	-0.28
		HI	12	37.92	30.26	8.74			
	T2	LO	8	31.25	27.87	9.85	-0.20	0.846	-0.09
		HI	12	33.75	27.81	8.03			
	T3	LO	8	29.38	29.69	10.50	-0.59	0.559	-0.27
		HI	12	36.67	24.89	7.19			
	T4	LO	8	35.00	31.74	11.22	-0.25	0.807	-0.11
		HI	12	38.33	27.99	8.08			
Temporal Demand	T1	LO	8	69.38	27.57	9.75	1.17	0.259	0.53
		HI	12	55.42	25.36	7.32			
	T2	LO	8	61.88	28.28	10.00	0.89	0.388	0.40
		HI	12	52.08	21.26	6.14			
	T3	LO	8	58.13	19.45	6.88	0.94	0.360	0.43
		HI	12	46.25	31.85	9.19			
	T4	LO	8	69.38	12.37	4.38	-0.21	0.835	-0.10
		HI	12	71.25	22.88	6.60			

Table L-34 Workload (NASA-TLX) statistics by high/low conscientiousness scores (continued)

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Effort	T1	LO	8	63.75	20.83	7.36	-0.47	0.643	-0.22
		HI	12	67.08	10.76	3.11			
	T2	LO	8	60.00	24.78	8.76	0.90	0.382	0.41
		HI	12	50.00	24.21	6.99			
	T3	LO	8	51.88	18.11	6.40	0.02	0.984	0.01
		HI	12	51.67	24.15	6.97			
	T4	LO	8	64.38	19.35	6.84	-0.40	0.691	-0.18
		HI	12	68.33	22.70	6.55			
Frustration	T1	LO	8	43.75	32.81	11.60	0.65	0.525	0.30
		HI	12	35.00	27.30	7.88			
	T2	LO	8	36.88	25.63	9.06	0.66	0.516	0.30
		HI	12	29.17	25.39	7.33			
	T3	LO	8	32.50	20.87	7.38	0.42	0.682	0.19
		HI	12	27.92	25.98	7.50			
	T4	LO	8	41.25	25.46	9.00	0.40	0.693	0.18
		HI	12	36.25	28.45	8.21			
Performance Satisfaction	T1	LO	8	30.00	17.73	6.27	-0.20	0.845	-0.09
		HI	12	31.67	18.87	5.45			
	T2	LO	8	40.00	18.90	6.68	1.63	0.122	0.74
		HI	12	26.67	17.36	5.01			
	T3	LO	8	30.63	14.50	5.13	1.37	0.187	0.63
		HI	12	21.67	14.20	4.10			
	T4	LO	8	42.50	20.00	7.07	0.22	0.825	0.10
		HI	12	40.00	26.88	7.76			

Table L-35 Workload (NASA-TLX) statistics by high/low neuroticism scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Overall Workload	T1	LO	10	49.90	17.24	5.45	-0.34	0.735	-0.15
		HI	10	52.50	16.55	5.24			
	T2	LO	10	43.30	17.14	5.42	-0.32	0.751	-0.14
		HI	10	45.90	18.85	5.96			
	T3	LO	10	41.90	18.16	5.74	0.12	0.908	0.05
		HI	10	41.00	16.24	5.14			
	T4	LO	10	52.70	19.37	6.13	-0.19	0.849	-0.09
		HI	10	54.20	15.09	4.77			
Mental Demand	T1	LO	10	76.50	16.51	5.22	0.40	0.694	0.18
		HI	10	73.50	17.00	5.38			
	T2	LO	10	59.00	25.58	8.09	-0.04	0.967	-0.02
		HI	10	59.50	26.92	8.51			
	T3	LO	10	55.50	25.22	7.97	0.09	0.931	0.04
		HI	10	54.50	25.54	8.08			
	T4	LO	10	68.50	26.15	8.27	0.23	0.823	0.10
		HI	10	66.00	23.07	7.30			
Physical Demand	T1	LO	10	39.00	31.61	9.99	0.68	0.503	0.31
		HI	10	30.50	23.39	7.40			
	T2	LO	10	34.50	29.01	9.17	0.28	0.782	0.13
		HI	10	31.00	26.54	8.39			
	T3	LO	10	36.00	27.87	8.81	0.37	0.714	0.17
		HI	10	31.50	26.15	8.27			
	T4	LO	10	38.50	31.89	10.08	0.23	0.823	0.10
		HI	10	35.50	26.92	8.51			
Temporal Demand	T1	LO	10	52.50	25.74	8.14	-1.48	0.156	-0.66
		HI	10	69.50	25.65	8.11			
	T2	LO	10	49.00	23.43	7.41	-1.32	0.202	-0.59
		HI	10	63.00	23.83	7.54			
	T3	LO	10	49.50	30.41	9.62	-0.24	0.815	-0.11
		HI	10	52.50	26.06	8.24			
	T4	LO	10	70.00	22.61	7.15	-0.11	0.910	-0.05
		HI	10	71.00	15.78	4.99			

Table L-35 Workload (NASA-TLX) statistics by high/low neuroticism scores (continued)

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Effort	T1	LO	10	61.00	14.30	4.52	-1.44	0.167	-0.64
		HI	10	70.50	15.17	4.80			
	T2	LO	10	53.50	20.55	6.50	-0.09	0.930	-0.04
		HI	10	54.50	28.72	9.08			
	T3	LO	10	51.00	23.31	7.37	-0.15	0.880	-0.07
		HI	10	52.50	20.58	6.51			
	T4	LO	10	67.00	22.75	7.20	0.05	0.959	0.02
		HI	10	66.50	20.28	6.41			
Frustration	T1	LO	10	36.00	25.91	8.19	-0.38	0.712	-0.17
		HI	10	41.00	33.23	10.51			
	T2	LO	10	28.00	22.51	7.12	-0.75	0.464	-0.33
		HI	10	36.50	27.99	8.85			
	T3	LO	10	27.00	25.52	8.07	-0.51	0.615	-0.23
		HI	10	32.50	22.52	7.12			
	T4	LO	10	27.00	26.48	8.37	-2.03	0.057	-0.91
		HI	10	49.50	22.91	7.24			
Performance Satisfaction	T1	LO	10	34.50	19.21	6.08	0.87	0.398	0.39
		HI	10	27.50	16.87	5.34			
	T2	LO	10	33.50	20.42	6.46	0.35	0.731	0.16
		HI	10	30.50	17.87	5.65			
	T3	LO	10	29.00	16.63	5.26	1.16	0.263	0.52
		HI	10	21.50	12.03	3.80			
	T4	LO	10	46.00	30.07	9.51	0.94	0.362	0.42
		HI	10	36.00	15.42	4.88			

Table L-36 Workload (NASA-TLX) statistics by high/low openness scores

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Overall Workload	T1	LO	7	50.14	22.08	8.35	-0.20	0.840	-0.10
		HI	13	51.77	13.66	3.79			
	T2	LO	7	45.71	22.34	8.44	0.20	0.842	0.09
		HI	13	44.00	15.46	4.29			
	T3	LO	7	44.86	18.22	6.88	0.66	0.520	0.31
		HI	13	39.62	16.41	4.55			
	T4	LO	7	58.00	23.57	8.91	0.88	0.392	0.41
		HI	13	51.00	12.53	3.48			
Mental Demand	T1	LO	7	68.57	19.52	7.38	-1.31	0.206	-0.62
		HI	13	78.46	14.05	3.90			
	T2	LO	7	47.14	27.97	10.57	-1.62	0.123	-0.76
		HI	13	65.77	22.62	6.27			
	T3	LO	7	50.00	25.50	9.64	-0.65	0.521	-0.31
		HI	13	57.69	24.88	6.90			
	T4	LO	7	60.71	28.49	10.77	-0.89	0.387	-0.42
		HI	13	70.77	21.68	6.01			
Physical Demand	T1	LO	7	46.43	31.32	11.84	1.44	0.168	0.67
		HI	13	28.46	24.01	6.66			
	T2	LO	7	49.29	31.68	11.97	2.19	0.042	1.03
		HI	13	23.85	20.43	5.67			
	T3	LO	7	47.86	32.26	12.19	1.86	0.079	0.87
		HI	13	26.15	20.12	5.58			
	T4	LO	7	52.14	40.81	15.42	1.83	0.084	0.86
		HI	13	28.85	16.48	4.57			
Temporal Demand	T1	LO	7	64.29	24.90	9.41	0.40	0.695	0.19
		HI	13	59.23	28.13	7.80			
	T2	LO	7	54.29	25.73	9.72	-0.23	0.823	-0.11
		HI	13	56.92	24.20	6.71			
	T3	LO	7	52.86	18.22	6.89	0.22	0.832	0.10
		HI	13	50.00	32.21	8.93			
	T4	LO	7	71.43	20.76	7.85	0.16	0.877	0.07
		HI	13	70.00	18.82	5.22			

Table L-36 Workload (NASA-TLX) statistics by high/low openness scores (continued)

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Effort	T1	LO	7	60.00	17.56	6.64	-1.26	0.222	-0.59
		HI	13	68.85	13.41	3.72			
	T2	LO	7	47.14	24.30	9.18	-0.92	0.369	-0.43
		HI	13	57.69	24.46	6.78			
	T3	LO	7	51.43	21.74	8.22	-0.05	0.962	-0.02
		HI	13	51.92	22.13	6.14			
	T4	LO	7	67.86	24.64	9.31	0.17	0.868	0.08
		HI	13	66.15	19.81	5.49			
Frustration	T1	LO	7	40.71	28.05	10.60	0.24	0.811	0.11
		HI	13	37.31	30.73	8.52			
	T2	LO	7	42.86	24.64	9.31	1.42	0.172	0.67
		HI	13	26.54	24.36	6.76			
	T3	LO	7	37.86	21.19	8.01	1.14	0.271	0.53
		HI	13	25.38	24.45	6.78			
	T4	LO	7	52.14	25.31	9.57	1.80	0.088	0.85
		HI	13	30.77	25.24	7.00			
Performance Satisfaction	T1	LO	7	28.57	11.07	4.19	-0.43	0.669	-0.20
		HI	13	32.31	21.08	5.85			
	T2	LO	7	32.86	13.80	5.22	0.15	0.885	0.07
		HI	13	31.54	21.45	5.95			
	T3	LO	7	28.57	10.29	3.89	0.74	0.472	0.34
		HI	13	23.46	16.63	4.61			
	T4	LO	7	43.57	17.49	6.61	0.35	0.733	0.16
		HI	13	39.62	27.19	7.54			

Table L-37 Situation awareness scores by high/low extroversion

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Situation Awareness Level 1	T1	LO	12	0.69	0.11	0.03	-0.31	0.757	-0.14
		HI	8	0.71	0.09	0.03			
	T2	LO	12	0.67	0.19	0.05	0.06	0.956	0.03
		HI	8	0.66	0.12	0.04			
	T3	LO	12	0.66	0.16	0.05	-0.39	0.700	-0.18
		HI	8	0.69	0.16	0.06			
	T4	LO	12	0.62	0.16	0.05	-0.97	0.347	-0.44
		HI	8	0.69	0.16	0.06			
Situation Awareness Level 2	T1	LO	12	0.33	0.20	0.06	0.00	1.000	0.00
		HI	8	0.33	0.18	0.06			
	T2	LO	12	0.46	0.30	0.09	0.74	0.471	0.34
		HI	8	0.38	0.12	0.04			
	T3	LO	12	0.47	0.23	0.07	-0.06	0.951	-0.03
		HI	8	0.48	0.26	0.09			
	T4	LO	12	0.44	0.29	0.08	0.60	0.558	0.27
		HI	8	0.38	0.19	0.07			
Perceived Reliability	T1	LO	12	1.21	0.38	0.11	0.85	0.409	0.39
		HI	8	1.06	0.37	0.13			
	T2	LO	12	1.29	0.32	0.09	1.59	0.130	0.72
		HI	8	1.09	0.19	0.07			
	T3	LO	12	1.38	0.39	0.11	0.90	0.381	0.41
		HI	8	1.22	0.36	0.13			
	T4	LO	12	1.21	0.54	0.16	1.03	0.316	0.47
		HI	8	0.97	0.45	0.16			

Table L-38 Situation awareness scores by high/low agreeableness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Situation Awareness Level 1	T1	LO	11	0.67	0.08	0.02	-1.25	0.227	-0.56
		HI	9	0.73	0.12	0.04			
	T2	LO	11	0.63	0.16	0.05	-1.18	0.254	-0.53
		HI	9	0.71	0.15	0.05			
	T3	LO	11	0.62	0.18	0.06	-1.68	0.109	-0.76
		HI	9	0.73	0.10	0.03			
	T4	LO	11	0.65	0.20	0.06	0.01	0.989	0.01
		HI	9	0.64	0.10	0.03			
Situation Awareness Level 2	T1	LO	11	0.29	0.18	0.06	-1.21	0.240	-0.55
		HI	9	0.39	0.19	0.06			
	T2	LO	11	0.41	0.25	0.08	-0.31	0.757	-0.14
		HI	9	0.44	0.25	0.08			
	T3	LO	11	0.41	0.20	0.06	-1.41	0.176	-0.63
		HI	9	0.56	0.26	0.09			
	T4	LO	11	0.39	0.24	0.07	-0.44	0.666	-0.20
		HI	9	0.44	0.28	0.09			
Perceived Reliability	T1	LO	11	1.07	0.46	0.14	-1.08	0.293	-0.49
		HI	9	1.25	0.22	0.07			
	T2	LO	11	1.20	0.27	0.08	-0.13	0.894	-0.06
		HI	9	1.22	0.32	0.11			
	T3	LO	11	1.16	0.26	0.08	-2.19	0.042	-0.98
		HI	9	1.50	0.43	0.14			
	T4	LO	11	1.02	0.47	0.14	-0.87	0.398	-0.39
		HI	9	1.22	0.57	0.19			

Table L-39 Situation awareness scores by high/low conscientiousness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Situation Awareness Level 1	T1	LO	11	0.68	0.10	0.03	-1.01	0.327	-0.45
		HI	9	0.72	0.10	0.03			
	T2	LO	11	0.61	0.18	0.05	-1.83	0.083	-0.82
		HI	9	0.73	0.11	0.04			
	T3	LO	11	0.63	0.15	0.04	-1.36	0.192	-0.61
		HI	9	0.72	0.16	0.05			
	T4	LO	11	0.61	0.18	0.05	-1.11	0.280	-0.50
		HI	9	0.69	0.13	0.04			
Situation Awareness Level 2	T1	LO	11	0.33	0.21	0.06	0.00	1.000	0.00
		HI	9	0.33	0.17	0.06			
	T2	LO	11	0.41	0.22	0.06	-0.31	0.757	-0.14
		HI	9	0.44	0.29	0.10			
	T3	LO	11	0.44	0.17	0.05	-0.73	0.473	-0.33
		HI	9	0.52	0.31	0.10			
	T4	LO	11	0.35	0.24	0.07	-1.38	0.186	-0.62
		HI	9	0.50	0.25	0.08			
Perceived Reliability	T1	LO	11	1.18	0.45	0.14	0.41	0.687	0.18
		HI	9	1.11	0.28	0.09			
	T2	LO	11	1.18	0.25	0.08	-0.52	0.607	-0.24
		HI	9	1.25	0.33	0.11			
	T3	LO	11	1.27	0.33	0.10	-0.51	0.618	-0.23
		HI	9	1.36	0.45	0.15			
	T4	LO	11	0.98	0.53	0.16	-1.34	0.197	-0.60
		HI	9	1.28	0.46	0.15			

Table L-40 Situation awareness scores by high/low neuroticism

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Situation Awareness Level 1	T1	LO	10	0.71	0.11	0.03	0.33	0.745	0.15
		HI	10	0.69	0.10	0.03			
	T2	LO	10	0.70	0.16	0.05	0.98	0.341	0.44
		HI	10	0.63	0.16	0.05			
	T3	LO	10	0.71	0.20	0.06	1.13	0.273	0.51
		HI	10	0.63	0.11	0.03			
	T4	LO	10	0.67	0.15	0.05	0.69	0.501	0.31
		HI	10	0.62	0.18	0.06			
Situation Awareness Level 2	T1	LO	10	0.32	0.17	0.05	-0.39	0.702	-0.17
		HI	10	0.35	0.21	0.07			
	T2	LO	10	0.47	0.26	0.08	0.75	0.461	0.34
		HI	10	0.38	0.24	0.07			
	T3	LO	10	0.52	0.30	0.09	0.78	0.447	0.35
		HI	10	0.43	0.16	0.05			
	T4	LO	10	0.48	0.29	0.09	1.20	0.245	0.54
		HI	10	0.35	0.20	0.06			
Perceived Reliability	T1	LO	10	1.10	0.27	0.08	-0.59	0.565	-0.26
		HI	10	1.20	0.47	0.15			
	T2	LO	10	1.18	0.31	0.10	-0.58	0.569	-0.26
		HI	10	1.25	0.26	0.08			
	T3	LO	10	1.33	0.44	0.14	0.14	0.887	0.06
		HI	10	1.30	0.33	0.10			
	T4	LO	10	1.23	0.55	0.17	0.99	0.337	0.44
		HI	10	1.00	0.47	0.15			

Table L-41 Situation awareness scores by high/low openness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Situation Awareness Level 1	T1	LO	10	0.66	0.10	0.03	-1.78	0.091	-0.80
		HI	10	0.74	0.08	0.03			
	T2	LO	10	0.64	0.13	0.04	-0.69	0.499	-0.31
		HI	10	0.69	0.19	0.06			
	T3	LO	10	0.67	0.13	0.04	0.00	1.000	0.00
		HI	10	0.67	0.19	0.06			
	T4	LO	10	0.62	0.20	0.06	-0.69	0.501	-0.31
		HI	10	0.67	0.12	0.04			
Situation Awareness Level 2	T1	LO	10	0.27	0.20	0.06	-1.66	0.113	-0.74
		HI	10	0.40	0.16	0.05			
	T2	LO	10	0.45	0.30	0.10	0.45	0.660	0.20
		HI	10	0.40	0.18	0.06			
	T3	LO	10	0.47	0.25	0.08	-0.15	0.880	-0.07
		HI	10	0.48	0.24	0.08			
	T4	LO	10	0.45	0.30	0.10	0.58	0.566	0.26
		HI	10	0.38	0.19	0.06			
Perceived Reliability	T1	LO	10	1.05	0.42	0.13	-1.21	0.243	-0.54
		HI	10	1.25	0.31	0.10			
	T2	LO	10	1.23	0.32	0.10	0.19	0.850	0.09
		HI	10	1.20	0.26	0.08			
	T3	LO	10	1.33	0.35	0.11	0.14	0.887	0.06
		HI	10	1.30	0.42	0.13			
	T4	LO	10	1.05	0.64	0.20	-0.54	0.597	-0.24
		HI	10	1.18	0.35	0.11			

Table L-42 Jian trust survey scores by high/low extroversion

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Trust in the Agent	T1	LO	9	54.33	11.39	3.80	-0.40	0.697	-0.18
		HI	11	56.64	14.10	4.25			
	T2	LO	9	56.33	15.60	5.20	-0.79	0.439	-0.36
		HI	11	60.55	7.62	2.30			
	T3	LO	9	59.33	13.29	4.43	-0.38	0.706	-0.17
		HI	11	61.36	10.41	3.14			
	T4	LO	9	55.56	13.88	4.63	0.71	0.489	0.32
		HI	11	52.27	6.20	1.87			

Table L-43 Jian trust survey scores by high/low agreeableness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Trust in the Agent	T1	LO	11	57.82	10.87	3.28	0.86	0.402	0.39
		HI	9	52.89	14.79	4.93			
	T2	LO	11	56.18	11.83	3.57	-1.04	0.311	-0.47
		HI	9	61.67	11.55	3.85			
	T3	LO	11	59.64	8.82	2.66	-0.34	0.737	-0.15
		HI	9	61.44	14.67	4.89			
	T4	LO	11	53.18	8.26	2.49	-0.27	0.791	-0.12
		HI	9	54.44	12.70	4.23			

Table L-44 Jian trust survey scores by high/low conscientiousness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Trust in the Agent	T1	LO	8	52.00	9.12	3.22	-1.04	0.312	-0.47
		HI	12	58.00	14.45	4.17			
	T2	LO	8	51.75	12.74	4.50	-2.40	0.027	-1.10
		HI	12	63.25	8.77	2.53			
	T3	LO	8	55.38	10.93	3.86	-1.69	0.109	-0.77
		HI	12	63.83	11.03	3.18			
	T4	LO	8	47.38	9.38	3.32	-2.60	0.018	-1.19
		HI	12	58.00	8.64	2.50			

Table L-45 Jian trust survey scores by high/low neuroticism

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Trust in the Agent	T1	LO	10	54.40	13.58	4.30	-0.41	0.684	-0.19
		HI	10	56.80	12.31	3.89			
	T2	LO	10	60.20	9.02	2.85	0.58	0.569	0.26
		HI	10	57.10	14.29	4.52			
	T3	LO	10	61.20	12.06	3.81	0.28	0.779	0.13
		HI	10	59.70	11.53	3.65			
	T4	LO	10	57.20	10.79	3.41	1.57	0.134	0.70
		HI	10	50.30	8.78	2.78			

Table L-46 Jian trust survey scores by high/low openness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Trust in the Agent	T1	LO	7	56.57	8.68	3.28	0.25	0.809	0.11
		HI	13	55.08	14.69	4.08			
	T2	LO	7	55.43	7.85	2.97	-0.90	0.382	-0.42
		HI	13	60.38	13.34	3.70			
	T3	LO	7	57.71	8.92	3.37	-0.77	0.450	-0.36
		HI	13	61.92	12.78	3.54			
	T4	LO	7	49.43	11.59	4.38	-1.43	0.171	-0.67
		HI	13	56.08	9.00	2.50			

Table L-47 Godspeed survey scores by high/low extroversion

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Anthropomorphism	T1	LO	9	2.67	0.60	0.20	-0.73	0.474	-0.33
		HI	11	2.87	0.65	0.19			
	T2	LO	9	2.73	0.82	0.27	-0.95	0.357	-0.43
		HI	11	3.07	0.78	0.23			
	T3	LO	9	2.58	0.82	0.27	-1.10	0.287	-0.49
		HI	11	3.04	1.01	0.30			
	T4	LO	9	2.38	0.72	0.24	-0.95	0.355	-0.43
		HI	11	2.67	0.67	0.20			
Animacy	T1	LO	9	2.82	0.74	0.25	-1.20	0.245	-0.54
		HI	11	3.24	0.83	0.25			
	T2	LO	9	2.91	0.97	0.32	-1.28	0.215	-0.58
		HI	11	3.38	0.66	0.20			
	T3	LO	9	3.06	1.07	0.36	-0.53	0.600	-0.24
		HI	11	3.29	0.87	0.26			
	T4	LO	9	2.74	0.73	0.24	-0.71	0.484	-0.32
		HI	11	3.01	0.94	0.28			
Likeability	T1	LO	9	3.60	0.66	0.22	0.45	0.658	0.20
		HI	11	3.45	0.76	0.23			
	T2	LO	9	3.47	0.77	0.26	-0.56	0.582	-0.25
		HI	11	3.64	0.58	0.17			
	T3	LO	9	3.40	0.87	0.29	-0.41	0.684	-0.19
		HI	11	3.56	0.88	0.27			
	T4	LO	9	3.13	0.70	0.23	-0.16	0.877	-0.07
		HI	11	3.18	0.68	0.20			
Perceived Intelligence	T1	LO	9	4.04	0.42	0.14	-0.28	0.782	-0.13
		HI	11	4.11	0.58	0.17			
	T2	LO	9	4.11	0.58	0.19	-0.08	0.941	-0.03
		HI	11	4.13	0.38	0.12			
	T3	LO	9	4.09	1.05	0.35	0.60	0.559	0.27
		HI	11	3.85	0.70	0.21			
	T4	LO	9	3.64	0.68	0.23	0.51	0.613	0.23
		HI	11	3.49	0.65	0.20			
Perceived Safety	T1	LO	9	3.45	0.55	0.18	0.51	0.616	0.23
		HI	11	3.30	0.67	0.20			
	T2	LO	9	3.26	0.49	0.16	-0.21	0.833	-0.10
		HI	11	3.30	0.38	0.11			
	T3	LO	9	3.37	0.45	0.15	-0.08	0.934	-0.04
		HI	11	3.40	0.80	0.24			
	T4	LO	9	3.19	0.41	0.14	-0.23	0.823	-0.10
		HI	11	3.23	0.47	0.14			

Table L-48 Godspeed survey scores by high/low agreeableness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Anthropomorphism	T1	LO	11	2.69	0.69	0.21	-0.70	0.491	-0.32
		HI	9	2.89	0.54	0.18			
	T2	LO	11	2.76	0.95	0.29	-0.97	0.345	-0.44
		HI	9	3.11	0.55	0.18			
	T3	LO	11	2.76	1.08	0.33	-0.34	0.736	-0.15
		HI	9	2.91	0.78	0.26			
	T4	LO	11	2.35	0.69	0.21	-1.43	0.169	-0.64
		HI	9	2.78	0.64	0.21			
Animacy	T1	LO	11	3.00	0.97	0.29	-0.30	0.768	-0.13
		HI	9	3.11	0.58	0.19			
	T2	LO	11	2.91	0.99	0.30	-1.61	0.126	-0.72
		HI	9	3.48	0.45	0.15			
	T3	LO	11	3.05	1.16	0.35	-0.71	0.485	-0.32
		HI	9	3.35	0.63	0.21			
	T4	LO	11	2.64	0.91	0.27	-1.55	0.139	-0.70
		HI	9	3.20	0.66	0.22			
Likeability	T1	LO	11	3.62	0.58	0.18	0.68	0.505	0.31
		HI	9	3.40	0.85	0.28			
	T2	LO	11	3.55	0.59	0.18	-0.11	0.917	-0.05
		HI	9	3.58	0.77	0.26			
	T3	LO	11	3.55	0.78	0.23	0.31	0.759	0.14
		HI	9	3.42	1.00	0.33			
	T4	LO	11	3.05	0.62	0.19	-0.77	0.451	-0.35
		HI	9	3.29	0.74	0.25			
Perceived Intelligence	T1	LO	11	4.07	0.49	0.15	-0.07	0.945	-0.03
		HI	9	4.09	0.54	0.18			
	T2	LO	11	4.02	0.49	0.15	-1.09	0.291	-0.49
		HI	9	4.24	0.42	0.14			
	T3	LO	11	3.91	0.90	0.27	-0.29	0.778	-0.13
		HI	9	4.02	0.85	0.28			
	T4	LO	11	3.44	0.59	0.18	-0.94	0.362	-0.42
		HI	9	3.71	0.72	0.24			
Perceived Safety	T1	LO	11	3.45	0.58	0.18	0.70	0.493	0.31
		HI	9	3.26	0.66	0.22			
	T2	LO	11	3.27	0.44	0.13	-0.12	0.907	-0.05
		HI	9	3.30	0.42	0.14			
	T3	LO	11	3.43	0.63	0.19	0.31	0.763	0.14
		HI	9	3.33	0.71	0.24			
	T4	LO	11	3.11	0.47	0.14	-1.16	0.262	-0.52
		HI	9	3.33	0.37	0.12			

Table L-49 Godspeed survey scores by high/low conscientiousness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Anthropomorphism	T1	LO	8	2.48	0.41	0.15	-1.92	0.070	-0.88
		HI	12	2.98	0.66	0.19			
	T2	LO	8	2.55	0.72	0.26	-1.79	0.090	-0.82
		HI	12	3.17	0.77	0.22			
	T3	LO	8	2.33	0.61	0.22	-2.15	0.045	-0.98
		HI	12	3.17	0.98	0.28			
	T4	LO	8	2.30	0.47	0.16	-1.29	0.213	-0.59
		HI	12	2.70	0.78	0.23			
Animacy	T1	LO	8	2.52	0.67	0.24	-2.83	0.011	-1.29
		HI	12	3.40	0.69	0.20			
	T2	LO	8	2.67	0.89	0.31	-2.49	0.023	-1.14
		HI	12	3.50	0.61	0.18			
	T3	LO	8	2.81	1.05	0.37	-1.47	0.159	-0.67
		HI	12	3.43	0.83	0.24			
	T4	LO	8	2.61	0.57	0.20	-1.26	0.222	-0.58
		HI	12	3.08	0.96	0.28			
Likeability	T1	LO	8	3.43	0.43	0.15	-0.48	0.635	-0.22
		HI	12	3.58	0.85	0.25			
	T2	LO	8	3.23	0.54	0.19	-1.99	0.062	-0.91
		HI	12	3.78	0.66	0.19			
	T3	LO	8	3.13	0.63	0.22	-1.61	0.124	-0.74
		HI	12	3.73	0.93	0.27			
	T4	LO	8	3.00	0.45	0.16	-0.87	0.398	-0.40
		HI	12	3.27	0.78	0.23			
Perceived Intelligence	T1	LO	8	3.90	0.32	0.11	-1.34	0.196	-0.61
		HI	12	4.20	0.57	0.17			
	T2	LO	8	3.95	0.48	0.17	-1.36	0.189	-0.62
		HI	12	4.23	0.44	0.13			
	T3	LO	8	3.83	0.99	0.35	-0.56	0.580	-0.26
		HI	12	4.05	0.80	0.23			
	T4	LO	8	3.33	0.62	0.22	-1.35	0.195	-0.61
		HI	12	3.72	0.65	0.19			
Perceived Safety	T1	LO	8	3.29	0.57	0.20	-0.44	0.663	-0.20
		HI	12	3.42	0.65	0.19			
	T2	LO	8	3.09	0.43	0.15	-1.82	0.085	-0.83
		HI	12	3.42	0.38	0.11			
	T3	LO	8	3.25	0.46	0.16	-0.74	0.468	-0.34
		HI	12	3.47	0.76	0.22			
	T4	LO	8	3.13	0.40	0.14	-0.70	0.492	-0.32
		HI	12	3.27	0.46	0.13			

Table L-50 Godspeed survey scores by high/low neuroticism

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Anthropomorphism	T1	LO	10	2.54	0.58	0.18	-1.84	0.082	-0.82
		HI	10	3.02	0.58	0.18			
	T2	LO	10	2.82	0.70	0.22	-0.55	0.588	-0.25
		HI	10	3.02	0.91	0.29			
	T3	LO	10	2.92	0.89	0.28	0.42	0.678	0.19
		HI	10	2.74	1.02	0.32			
	T4	LO	10	2.50	0.74	0.24	-0.25	0.803	-0.11
		HI	10	2.58	0.67	0.21			
Animacy	T1	LO	10	2.85	0.64	0.20	-1.12	0.276	-0.50
		HI	10	3.25	0.92	0.29			
	T2	LO	10	3.17	0.76	0.24	0.00	1.000	0.00
		HI	10	3.17	0.93	0.29			
	T3	LO	10	3.12	0.89	0.28	-0.31	0.761	-0.14
		HI	10	3.25	1.05	0.33			
	T4	LO	10	2.85	0.90	0.28	-0.22	0.832	-0.10
		HI	10	2.93	0.83	0.26			
Likeability	T1	LO	10	3.36	0.72	0.23	-1.02	0.322	-0.46
		HI	10	3.68	0.68	0.22			
	T2	LO	10	3.64	0.61	0.19	0.53	0.602	0.24
		HI	10	3.48	0.73	0.23			
	T3	LO	10	3.56	0.95	0.30	0.36	0.726	0.16
		HI	10	3.42	0.81	0.26			
	T4	LO	10	3.22	0.75	0.24	0.39	0.700	0.18
		HI	10	3.10	0.61	0.19			
Perceived Intelligence	T1	LO	10	3.96	0.48	0.15	-1.08	0.295	-0.48
		HI	10	4.20	0.52	0.16			
	T2	LO	10	4.18	0.39	0.12	0.57	0.578	0.25
		HI	10	4.06	0.54	0.17			
	T3	LO	10	3.86	0.85	0.27	-0.51	0.616	-0.23
		HI	10	4.06	0.90	0.29			
	T4	LO	10	3.76	0.61	0.19	1.41	0.176	0.63
		HI	10	3.36	0.66	0.21			
Perceived Safety	T1	LO	10	3.17	0.61	0.19	-1.51	0.149	-0.67
		HI	10	3.57	0.57	0.18			
	T2	LO	10	3.43	0.32	0.10	1.65	0.117	0.74
		HI	10	3.14	0.48	0.15			
	T3	LO	10	3.60	0.70	0.22	1.55	0.139	0.69
		HI	10	3.17	0.55	0.17			
	T4	LO	10	3.27	0.41	0.13	0.57	0.573	0.26
		HI	10	3.15	0.47	0.15			

Table L-51 Godspeed survey scores by high/low openness

			<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>t</i> (18)	<i>p</i>	<i>d_s</i>
Anthropomorphism	T1	LO	7	2.74	0.49	0.18	-0.19	0.850	-0.09
		HI	13	2.80	0.70	0.19			
	T2	LO	7	2.97	0.51	0.19	0.21	0.839	0.10
		HI	13	2.89	0.93	0.26			
	T3	LO	7	2.74	0.34	0.13	-0.30	0.769	-0.14
		HI	13	2.88	1.15	0.32			
	T4	LO	7	2.40	0.43	0.16	-0.66	0.520	-0.31
		HI	13	2.62	0.80	0.22			
Animacy	T1	LO	7	3.00	0.68	0.26	-0.20	0.843	-0.09
		HI	13	3.08	0.88	0.24			
	T2	LO	7	3.19	0.41	0.16	0.09	0.926	0.04
		HI	13	3.15	1.00	0.28			
	T3	LO	7	3.24	0.46	0.17	0.18	0.860	0.08
		HI	13	3.16	1.15	0.32			
	T4	LO	7	2.71	0.48	0.18	-0.68	0.504	-0.32
		HI	13	2.99	0.99	0.27			
Likeability	T1	LO	7	3.34	0.30	0.11	-0.82	0.423	-0.38
		HI	13	3.62	0.84	0.23			
	T2	LO	7	3.23	0.44	0.17	-1.73	0.101	-0.81
		HI	13	3.74	0.70	0.20			
	T3	LO	7	3.29	0.50	0.19	-0.77	0.450	-0.36
		HI	13	3.60	1.00	0.28			
	T4	LO	7	2.86	0.22	0.08	-1.54	0.142	-0.72
		HI	13	3.32	0.78	0.22			
Perceived Intelligence	T1	LO	7	4.03	0.24	0.09	-0.33	0.746	-0.15
		HI	13	4.11	0.60	0.17			
	T2	LO	7	4.09	0.23	0.09	-0.24	0.816	-0.11
		HI	13	4.14	0.56	0.16			
	T3	LO	7	4.23	0.51	0.19	1.03	0.318	0.48
		HI	13	3.82	0.99	0.27			
	T4	LO	7	3.54	0.57	0.22	-0.08	0.934	-0.04
		HI	13	3.57	0.71	0.20			
Perceived Safety	T1	LO	7	3.43	0.37	0.14	0.32	0.750	0.15
		HI	13	3.33	0.72	0.20			
	T2	LO	7	3.19	0.42	0.16	-0.71	0.487	-0.33
		HI	13	3.33	0.43	0.12			
	T3	LO	7	3.24	0.46	0.17	-0.72	0.484	-0.34
		HI	13	3.46	0.74	0.21			
	T4	LO	7	3.00	0.38	0.15	-1.66	0.115	-0.78
		HI	13	3.32	0.43	0.12			

List of Symbols, Abbreviations, and Acronyms

A	agreeableness
AI	artificial intelligence
AiTR	Aided Target Recognition
ANOVA	analysis of variance
C	conscientiousness
E	extroversion
GQS	Godspeed Questionnaire Series
HAT	human-agent team/teaming
HTL	high task load
IA	inhibitory anxiety
ID	individual difference
IPIP	International Personality Item Pool
IUS	intolerance of uncertainty
LCA	learning-capable agent
LTL	low task load
M	mean
MD	mental demand
ML	machine learning
N	neuroticism
NASA	National Aeronautics and Space Administration
NGCV	Next Generation Combat Vehicle
O	openness
PA	prospective anxiety
PAC	perceived attentional control
PD	physical demand
RCV	robotic combat vehicle

s	seconds
SA	situation awareness
SAGAT	Situation Awareness Global Assessment Technique
SD	standard deviation
SE	standard error
TD	temporal demand
TLX	Task Load Index
UCF	University of Central Florida

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DEVCOM ARL
(PDF) FCDD RLB CI
TECH LIB

17 DEVCOM ARL
(PDF) FCDD RLA A
P FRANASZCZUK
FCDD RLA F
J GASTON
Y-S CHEN
FCDD RLA FA
A MARATHE
FCDD RLA FB
JO GARCIA
FCDD RLA FC
J TOURYAN
H ROY
H CROWELL
FCDD RLA FD
KR COX
A BAKER
E HOLDER
SG LAKHMANI
AS KRAUSMAN
A SCHARINE
JL WRIGHT
J REXWINKLE
FCDD RLA FE
K MCDOWELL