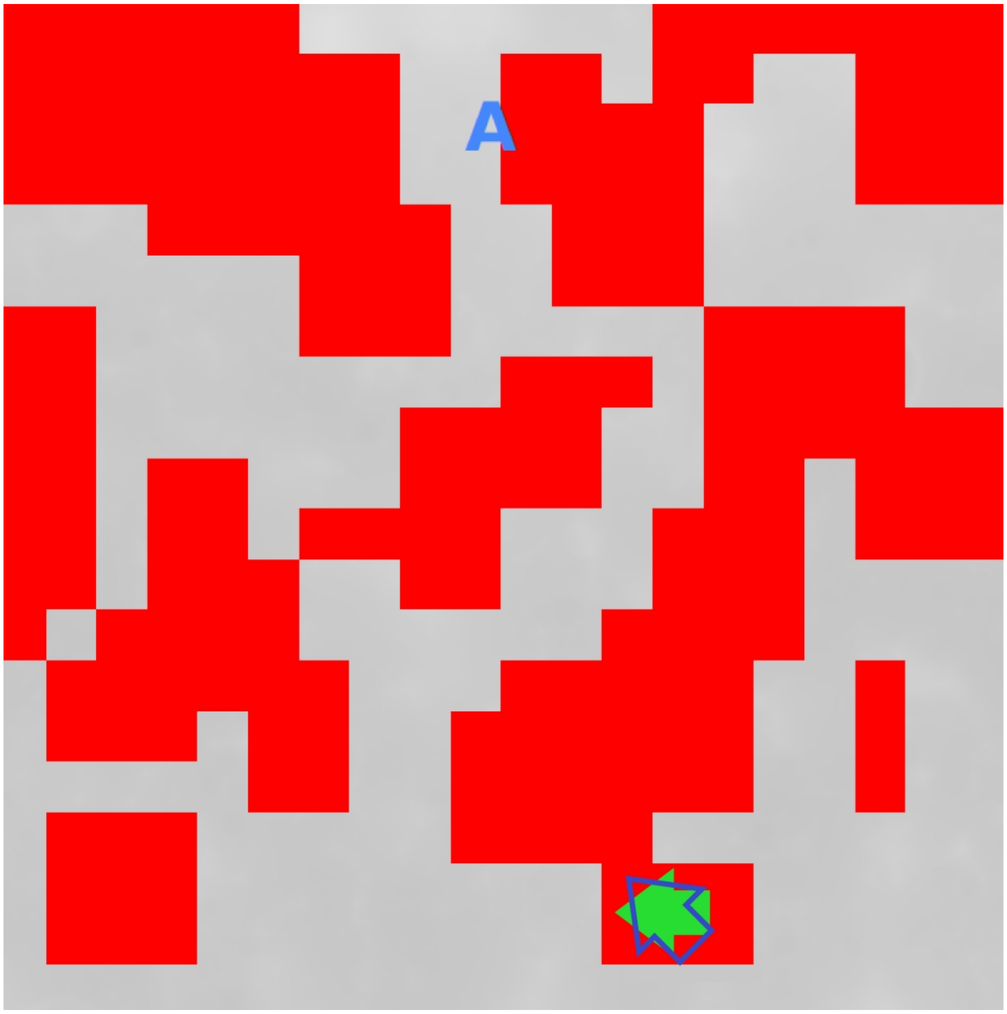
Human Factors

Human Operator Trust is Similar between Human versus Autonomous Navigational Assistance

Journal:	*Human Factors : The Journal of the Human Factors and Ergonomics Society*
Manuscript ID	Draft
Manuscript Type:	Research Article
Keywords:	Automation < AVIATION AND AEROSPACE, Human-automation interaction < AUTOMATION, EXPERT SYSTEMS, Trust in automation < AUTOMATION, EXPERT SYSTEMS, HUMAN-COMPUTER INTERACTION, COMPUTER SYSTEMS, HUMAN-ROBOT INTERACTION, Teleoperation < HUMAN-ROBOT INTERACTION


82%


A

Human Operator Trust is Similar between Human versus Autonomous Navigational Assistance

Jamie L Voros*, Jamison McGinley*, Steve McGuire*,‡, Michael E Walker†, Priyanka Karki*, Nisar Ahmed*, Daniel Szafir†, Torin K Clark*

* Ann & H.J. Smead Department of Aerospace Engineering Sciences, The University of Colorado Boulder

† Department of Computer Science, ATLAS Institute, The University of Colorado Boulder

‡ Department of Electrical and Computer Engineering, University of California Santa Cruz

Type: Research Article

Word count: 4418

Corresponding author: Jamie L Voros (jamie.voros@colorado.edu)

Acknowledgements

This work was supported by the University of Colorado Boulder Autonomous Systems Integrated Research Thrust (ASIRT) and an Early Career Faculty grant from NASA's Space Technology Research Grants Program under award NNX16AR58G.

Abstract

Objective: We examined how human operator trust in navigational assistance differed when the assistance was human vs autonomous.

Background: As autonomy becomes ever more ubiquitous, it is critical to understand how trust in autonomous systems differs from that in another human.

Method: Benign navigational assistance was provided by either another human or an autonomous system and presented in an identical manner. Half of the subjects were deceived and told the assistance was provided by the opposite source. We quantified trust as how closely subjects' rover driving actions aligned with recommendations given by the navigational assistant. This metric of trust is objective, continuous, and unobtrusive.

Results: The presence of the navigational assistance changed subject behavior ($p = 0.002$) but there was not a significant difference between trust in the human and autonomous navigational assistance systems. This suggests that our subject pool was not more or less trusting in an autonomous system, as compared to assistance from another human, particularly when controlling for the system's efficacy. Subjects self-reported their trust in the system after the experiment using a standard trust questionnaire. Self-reported trust correlated with objectively measured trust on difficult rover operating scenarios ($p = 0.01$, $r = 0.45$).

Conclusion: Our findings inform future human-autonomy teaming design choices and provide a unique approach to quantify operator trust.

Application: Potential applications include crewed deep space missions where communication delays may require ground controllers to be replaced with onboard autonomous systems while maintaining and quantifying trust throughout.

Keywords

Trust in automation, Human–automation interaction, Automation, Human-Computer Interaction, Human-robot interaction, Teleoperation

Précis

The rise of autonomous systems requires understanding of how we interact with them. In this study, we examined if human operators trust an autonomous versus human guidance system differently. We used two measures of operator trust: a traditional self-reported trust questionnaire and an unobtrusive, passive objective metric of trust.

Introduction

Suitably quantifying trust in an autonomous system is important because it has a significant impact on how (or if) the system is used (Holden et al., 2013; Sheridan, 2002). Having too much trust in a system can result in over-reliance (Parasuraman & Riley, 1997). Research on trust in an autonomous systems can be broken down into three parts: time invariant propensity to trust, time variant trust and instantaneous trust (Schaefer et al., 2016). Time invariant propensity to trust influences overall trust in a system; it accounts for 52% of trust variance (Merritt & Ilgen, 2008). Instantaneous trust can be volatile and is responsive to immediate environmental and situation conditions (Schaefer et al., 2016). An example of instantaneous trust is trust as measured by polling a subject during the experiment and an example of time variant trust might be the subject's overall trust at the conclusion of the experiment.

Time variant trust in both autonomous and human systems builds up on good system performance and declines with poor performance (Desai et al., 2013; Lewandowsky et al., 2000; Yang et al., 2017). Perceived competence of the system correlates with trust and observed system incompetency reduces trust (Hancock et al., 2011; Muir & Moray, 1996). Yang et al. (2017) presented a Bayesian model for linking subjects' instantaneous trust to their time variant trust. Ma and Kaber (2007) showed time changing trust behavior exists for trust in autonomous systems specifically for navigational assistance. They found that trust in an autonomous navigational assistant changed with system performance but did not change in the same way as it did for a human navigational assistant. They showed that while humans appeared to have more initial trust in the autonomous system, trust declined more sharply in the autonomous system when it made errors (Ma & Kaber, 2007).

Trust behavior (objective measures of trust) with respect to autonomous systems is a measure of the way humans use the system (Bindewald et al., 2018; Chiou & Lee, 2015; Dixon & Wickens, 2006; Lacson et al., 2005). Trust behaviors can be broken down into compliance, reliance, human override and human correction to an autonomous system (Bindewald et al., 2018). Muir and Moray (1996) showed that subjective trust in automation correlates with use of the system. Dzindolet et al. (2003) found that self-reported trust tends to predict reliance in an automated system aid. Although self-reported trust correlates with system use, self-reported trust in the system did not necessarily co-vary with implicit trust (Merritt et al., 2013). While trust can predict use of the system, studies do not always find significant correlation between trust and performance (Merritt et al., 2015; Voros et al., 2020). Dzindolet et al. (2002) showed that, despite equal performance of human and automated assistance, subjects were more likely to disuse the automated aid than human aid.

Changing the presentation of automated assistance changes trust in the system, even if the same information is being conveyed; more human-like information presentation results in higher trust (Li & Yeh, 2010; Nass & Moon, 2000; Stedmon et al., 2007). Since we were concerned with subjects' behavior in relation to (perceived) true origin of the assistance, it was important that the presentation was held constant across all groups.

Developing a framework for human-autonomy trust in space exploration is essential for long duration crewed missions (McGuire et al., 2018; NASA, 2019) . NASA's Human Research Roadmap identifies the lack of overall human-robotic system performance for space operations as a gap requiring investigation (NASA, 2019). Here, we aimed to investigate trust in an autonomous system, as compared to that in a human assistant specifically on a task controlling a lunar rover.

We quantify trust objectively by how closely aligned subjects' control inputs are to the recommendations of the navigational assistance. With this in mind, we ask two scientific questions: 1) How does the perceived source of navigational assistance (human versus autonomous) affect operator trust? 2) How does an objective measure of trust (alignment of subjects' control inputs to recommendations of the navigational assistance) compare to a standard questionnaire measure (Jian et al., 2000)?

Methods

Subjects drove a simulated lunar rover with limited battery to 'goals', while attempting to avoid terrain obstacles. Benign navigational assistance was provided in real-time by either a human or an autonomous system (except in the control condition, where no assistance was provided). In a between-subjects design, half of the subjects were deceivingly told the assistance was from the opposite source. We quantified trust in the navigational assistant objectively by comparing alignment of the rover's direction of travel to that suggested. We quantified overall self-reported trust in the system via Jian et al.'s (2000) "System Trust Scale" questionnaire.

Task

Figure 1 shows a screenshot of the subjects' view, displayed on a monitor. Subjects drove the simulated rover via a thumb stick, which specified the desired direction of travel. The rover had simulated dynamics such that it took a brief time to accelerate, decelerate, and change direction.

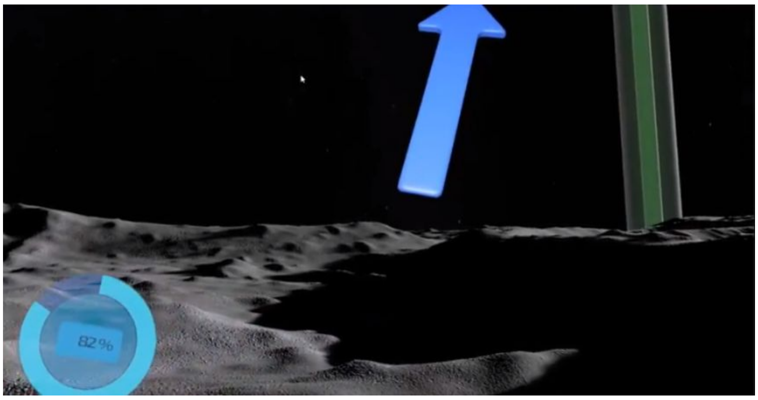


Figure 1: Screenshot of Subject View. The green beam denoted the goal, the blue arrow was the recommended direction of travel from the navigational assistant and the gauge in the bottom left showed the percentage of battery remaining.

The task had a series of six pre-set end 'goal' locations and associated pre-set start locations. The goal location was shown by a vertical green beam, such that it could be visible if behind a hill (Figure

1). Each ‘goal’ attempt from its corresponding ‘start’ was called a ‘traverse’. Subjects had to drive the simulated lunar rover across terrain littered with hazards without getting it stuck or running out of battery. If subjects were unable to reach a ‘goal’ location, that traverse would be recorded as ‘failed’ and the rover would automatically be transported to the following start location. Each traverse started with the rover facing the goal location, such that the green beam was within the camera view, and with the battery level at 100%. Subjects attempted the same six unique traverses in the same order. Traverses were of varying difficulty levels, though each could be completed with the allocated battery. Subjects were not provided more global information regarding the location of hazards farther away.

Navigational Assistance

To help the subject, two different types of navigational assistance were provided in this experiment; one human and one autonomous.

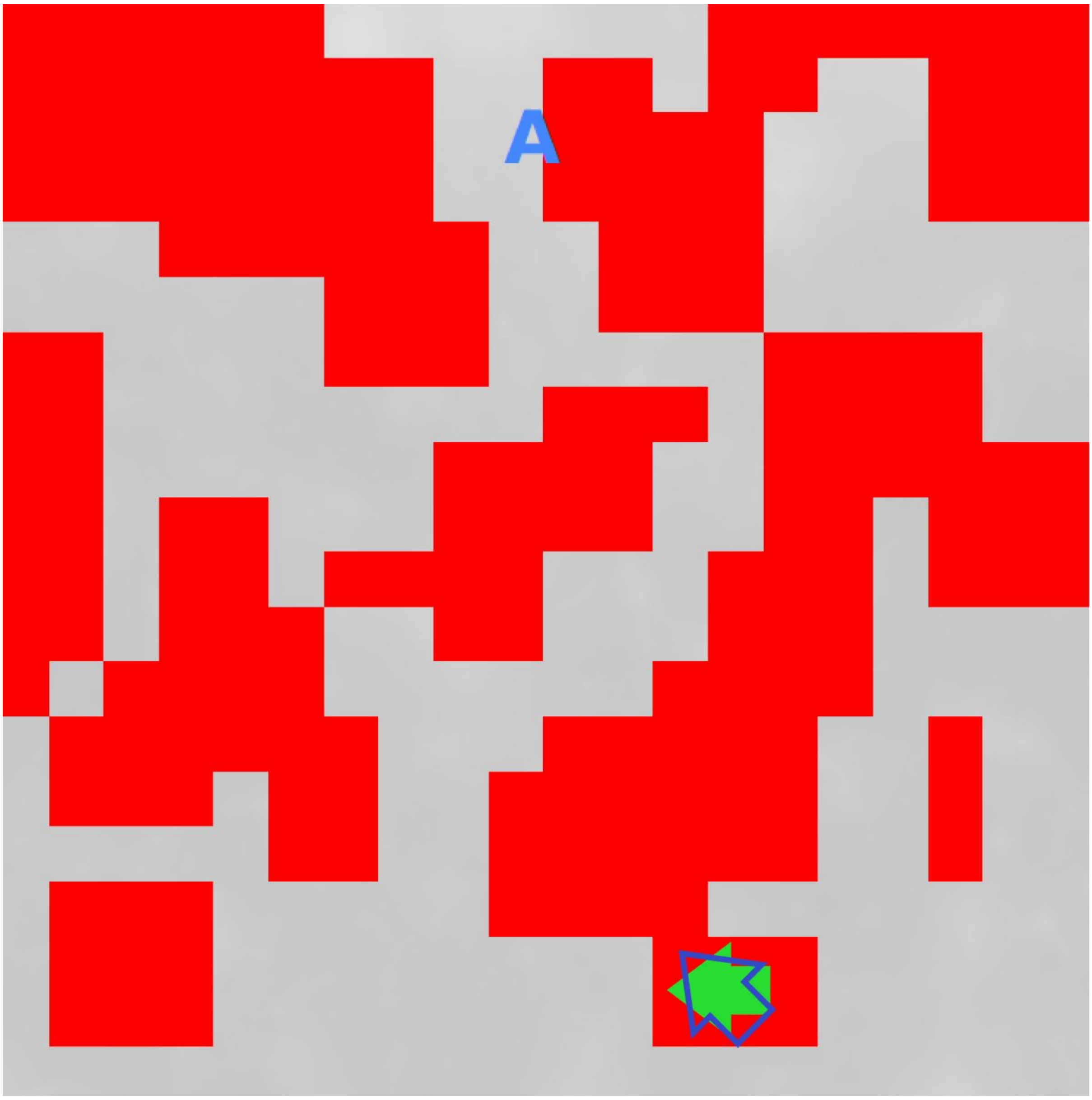


Figure 2: Hazard Map presented to both Human and Autonomous Guidance. Red indicated terrain that was more difficult to traverse. Green arrow indicated current recommended direction of travel. Blue arrow indicated location and direction of movement of rover. The blue letter 'A' was the goal location.

Critically, both types of guidance assistance were provided the same input information: the birds-eye view hazard map shown in Figure 2. The map depicted the lunar terrain broken down into lower resolution blocks that were categorized based upon whether that area was hazardous. The grey blocks indicated easier-to-cross terrain the red blocks indicated harder-to-cross terrain. The hazard categorization was done based upon the terrain's variation of height within the block. Grid blocks including terrain such as cliffs or steep slopes were characterized as hazardous, while generally flat terrain was characterized as not hazardous. The reduced resolution hazard map was intended to be representative of scenarios where the ground control may have more global, but less local information, complementing the crew who have the local information.

The blue arrow on the hazard map indicated the rover's current location and direction of travel and the green arrow was the current recommended direction of travel. The end goal location for the current traverse was denoted with a blue capital letter.

Autonomous Guidance System

The autonomous guidance system used a Markov Decision Process (MDP) to calculate the lowest weighted route to the block containing the goal (Kochenderfer, 2015). Red (harder-to-cross) blocks were assigned a higher weight than grey blocks (100x higher) to deter traversing hazardous terrain, even if it was the most direct route. Prior to the experiment, the MDP was solved for each goal to determine the optimal direction of travel when the rover was in each block. This was done by allowing for the direction of travel from each block to its neighboring blocks to be in each of the eight cardinal directions (north, northeast, east, southeast, south, southwest, west, northwest). Transition probabilities in the MDP formulation were set to have a 92% likelihood of the desired action (e.g., drive north) producing the expected transition (moving one block to the north of the current block), while all other transitions (moving one block to the northeast, east, etc., plus staying in the same block) each had a 1% likelihood.

We used an MDP for the autonomous guidance to provide a policy that maximizes an expected total reward function (i.e., penalized traversing hazardous terrain and rewarded reaching the goal efficiently). Our objective was not to evaluate the efficacy of the MDP, but instead to produce guidance from an automated system such that the responses could be compared to those receiving human assistance. During the experiment, the guidance direction dynamically updated as the subject drove the rover.

Human Guidance System

The human guidance system was a human research assistant. The research assistant was located in a neighboring room and was never seen by the subject to avoid biasing the subject's trust. During each traverse, the human research assistant looked at the hazard map (Figure 2) in real time and would input a suggested direction of travel via a thumb-stick. When the human research assistant updated their recommended direction of travel, the arrow displayed to the subject updated in real-time. The same human research assistant provided the human guidance throughout the study and received training on their task.

The guidance systems were both intended to be helpful and provide navigational assistance to the best of their abilities with the limited information given. The way the navigational assistance was displayed to the subject was identical in both cases so that the subject would be unable to distinguish what source of guidance they had.

Protocol

This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at The University of Colorado Boulder under protocol #18=0338. Informed consent was obtained from each participant. Thirty-nine subjects completed testing, of which 32 subjects received navigational guidance (Table 1) and seven were control subjects. Our subjects were recruited locally and largely consisted of undergraduate and graduate engineering students.

Subjects were randomly assigned the type of navigational guidance they would receive (autonomous, human or control). Control subjects received no guidance. Subjects not in the control group were randomly assigned the type of navigational assistance they were told they had (autonomous or human). Half of the guided subjects were deceived.

All subjects received an identical briefing on the navigational task, with the only variation in describing the source of navigational assistance they would receive. Subjects were told that there were other testing groups, but these groups were not explicitly indicated to help conceal the potential presence of deception.

Subjects completed at least one practice traverse before testing began. The practice provided an opportunity for subjects to familiarize themselves with the rover's driving dynamics and capabilities traversing terrain hazards. If the subject felt uncomfortable with their ability, or failed the practice task, they would attempt up to three unique practice traverses. The navigational guidance was disabled during practice in order to prevent premature interactions that could affect the subjects' subsequent use of the guidance. After practice, guided subjects were shown a video of the lunar rover in motion in order to become acquainted with the appearance and general behavior of the navigational guidance.

After practice, the experiment operator left the room and remotely initiated the experiment. The subject then drove the rover through a series of six lunar traverses in the same predetermined order. After the sixth traverse, the subject completed a post-experiment questionnaire, including the 12-question System Trust Scale (Jian et al., 2000). Control subjects did not complete the questionnaire since they did not interact with any navigational assistance. If a subject was assigned to a group in which they were deceived in the initial experimental briefing, after completion of the post-experiment questionnaire, they were read an explicit deception notice.

Independent variables

The two variables we manipulated were the type of guidance a subject had (autonomous, human or control) and the type of guidance a subject was told they had (autonomous, human or control). This 2x2 full-factorial experimental design, plus the control group, produced five total experimental groups: 'told autonomous, received autonomous' [tArA], 'told autonomous, received human' [tArH], 'told human, received autonomous' [tHrA], 'told human, received human' [tHrH] and 'control' [C]. While the subjects in the control group received no guidance, we did have the guidance human research assistant produce guidance recommendations in real-time. The sequence of guidance suggestions were recorded (but not displayed), such that they could be compared to the subject's route. The autonomous navigational assistance was calculated off-line based upon the rover's position for each control subject. Thus the control condition served as a baseline for behavior and performance when no guidance assistance was provided.

Table 1: Number of subjects per experimental group, defined by what source of guidance they were told and what was the actual source of navigational assistance. In addition, seven control [C] subjects were tested.

Actual Told	Autonomous (A)	Human (H)
Autonomous (A)	8 [tArA]	8 [tArH]
Human (H)	8 [tHrA]	8 [tHrH]

Dependent Variables

The angle between the rover's direction of travel and the recommended direction of travel was used as a measure of objective trust. Let s be the time-varying unit vector pointing in the rover's horizontal direction of travel and g be that denoting the recommended direction of travel.

Objectively measured trust, C, was computed as follows (bar indicative of average over time throughout each traverse):

$$C = \overline{\max\left(\frac{s \cdot g}{|s||g|}, 0\right)}$$

C values closer to one correspond to driving the rover in nearly the same direction recommended by the navigational assistance most of the time. Lower values were produced when the subject drove the rover in a very different direction than recommended (i.e., noncompliance), suggesting less trust in the navigational assistant. If the dot product between s and g was less than zero, these values were brought to zero because an angle discrepancy outside of ±90 degrees was universally classed as complete noncompliance. C_{all} was trust computed over the entire duration of all six traverses.

Objective trust in the control group was computed for both guidance systems. Each control subject had two data points per traverse (and overall) for objective trust, creating two more experimental groups: ‘control as compared to autonomous’ [CcA] and ‘control as compared to human’ [CcH].

Self-reported trust was collected via the post-experiment questionnaire, yielding one measure per subject (excluding control subjects). Table 2 summarizes the independent and dependent variables.

Table 2: Independent and dependent variables.

Independent Variables	**Dependent Variables**
Guidance Source: The guidance source the subject was actually provided. Human, autonomous or control (none).	**Objective Trust**: Computed as an average per traverse, and also overall.
Perceived Guidance Source: The guidance source the subject was told. Human, autonomous or control (none).	**Self-reported Trust**: Reported overall, integer response to series of 12 trust questions (Jian et al., 2000).

Data Processing

The post-experiment questionnaire included: “What kind of navigational assistant did you have for the experiment?” Our original intention was primarily to ensure that subjects recalled which source of guidance they were told. However, when we examined responses, we found a substantial portion of subjects were suspicious of the instructions; they did not believe that they had the type of guidance they were told. In particular, many subjects did not believe that there was a real human navigational assistant, even when not deceived [tHrH]. Since the source of the navigational assistance they were told was a primary independent variable, we removed subjects that provided any indication in their questionnaire response that the type of guidance they thought they had was not consistent with what they were told. For example, subject 1 wrote “I think it changed part of the way through” and was removed. After the removal of the suspicious subjects, very few subjects remained in either of the told human groups (tHrH or tHrA, see Table 3). Therefore, the two told human groups (shaded in Table 2) were pooled into one ‘told human and believed it’ [tH] group. The combined ‘told human and believed it’ (H) group ended up including a total of six subjects. The ‘told autonomous, received autonomous’ [tArA] and ‘told autonomous, received human’ [tArH] had seven and eight subjects respectively (i.e., only one subject in the tArA group was suspicious), and were analysed independently.

Table 3: Number of subjects per experimental group that believe the source of guidance they were told. Shaded groups were pooled into one ‘told human and believed it’ [H] group.

Actual Told	Autonomous (A)	Human (H)
Autonomous (A)	7 [tArA]	8 [tArH]
Human (H)	2 [tH]	4 [tH]

As our primary objective, we analyzed whether subjects' trust in the navigational guidance differed based upon whether its source was another human or an autonomous system. To control for the quality of the guidance being provided, we compared the 'told autonomous, received human' [tArH] and 'told autonomous, received autonomous' [tArA] groups.

In order to examine the impact of task difficulty on trust (traverses were designed to be of varying degrees of difficulty), traverses were classified as 'easy' or 'hard' depending on whether the majority of subjects completed them (i.e., >50% or <50% completion rate). Two traverses were classified as 'hard' (majority failed) and four as 'easy' (majority completed). Table 4 shows a breakdown of traverses by completion

Table 4: Breakdown of Traverses into 'Easy' or 'Difficult'

Traverse	**Completion Rate**	**Difficulty Classification**
A	0.67	Easy
B	0.90	Easy
C	0.82	Easy
D	0.21	Hard
E	0.36	Hard
F	0.56	Easy

When examining the relationship between objectively measured trust and self-reported trust, data from all subjects was used. We were interested in only the relationship between objective and self-reported trust and not the effect of guidance source.

Results

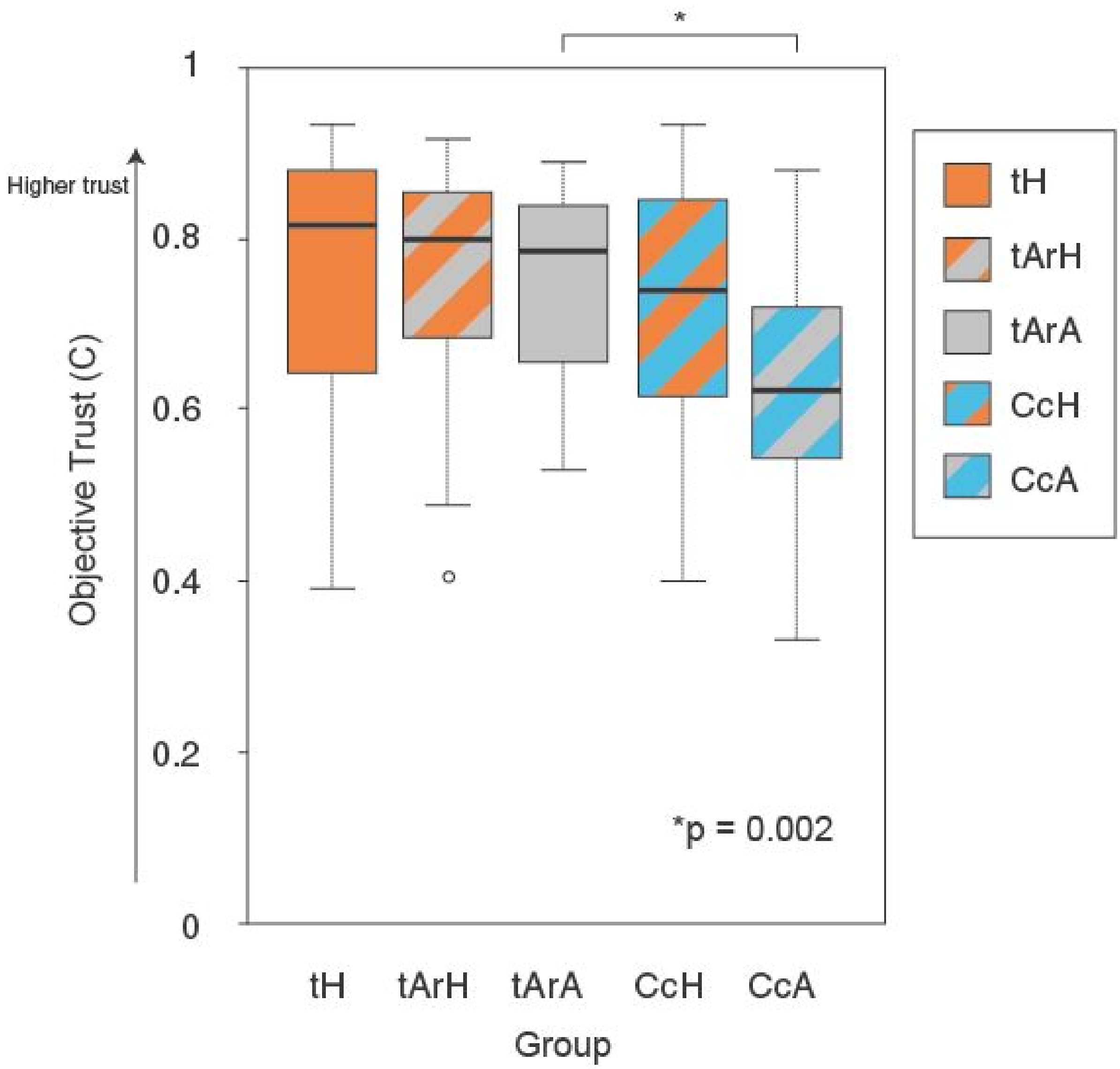
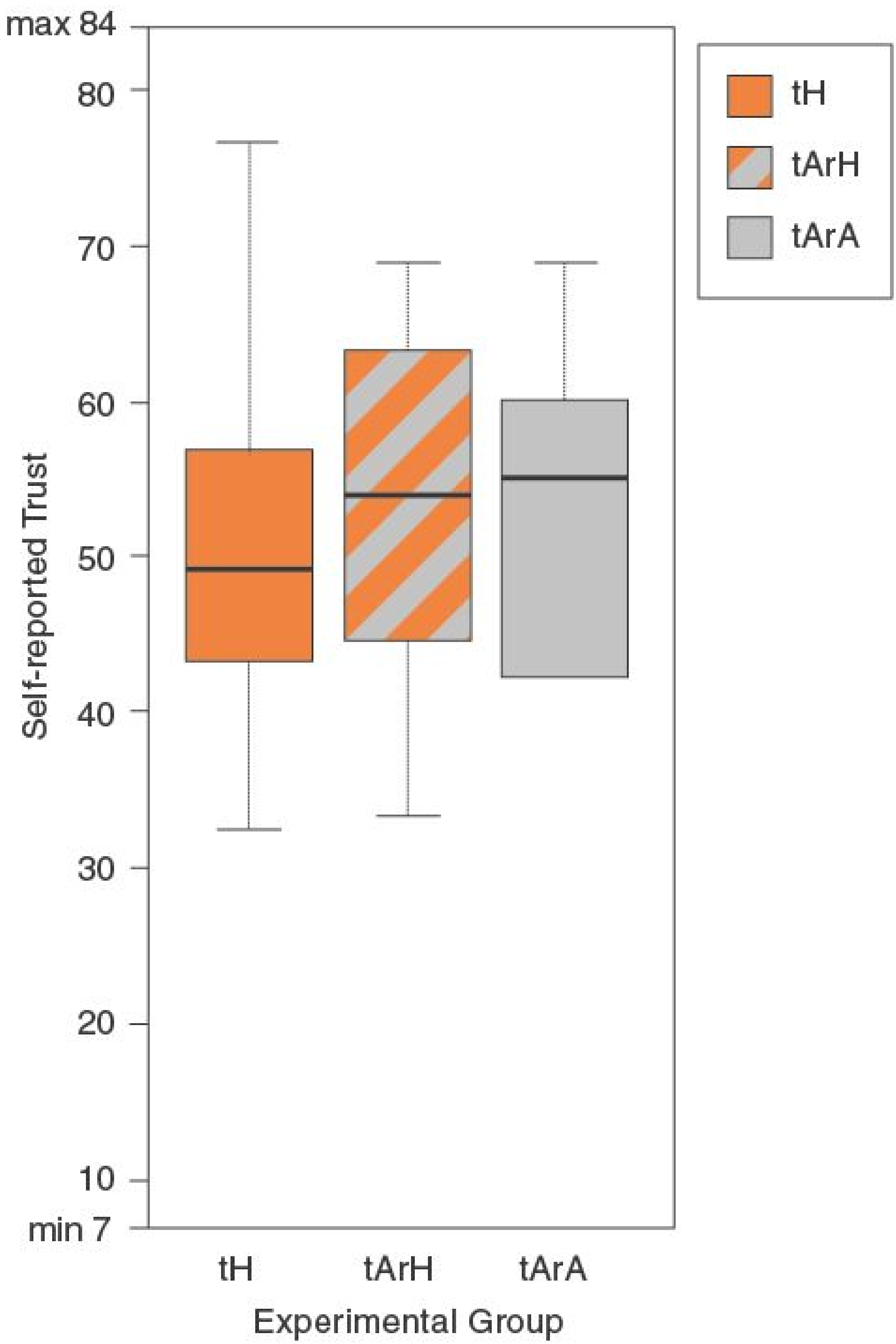


Figure 3 Boxplot to show differences in objective trust across experimental groups.

In comparing the objective trust metric across experimental groups, we found a significant difference between the control group as compared to autonomous guidance [CcA] and the told autonomous, received autonomous group [tArA] (Welch t-test for unequal samples, $p = 0.002$, $t(8.99) = 4.37$, 95% confidence interval (CI) = [0.05 - 0.17], mean difference = 0.11). This indicates that the presence of autonomous guidance changed subject behavior. There was no significant difference between the tArH and CcH groups (Welch t-test for unequal samples, $p = 0.49$, $t(12.66) = 0.10$, 95% CI = [-0.11, 0.06], mean difference = 0.028). Although sample sizes differed per group, the differences were small for the three guided groups ($n = 6$, 7 and 8), so we proceeded to use a one-way ANOVA to test for differences between the three groups with different guidance systems (tH, tArH, tArA). This found no significant difference in objective trust between the three guided groups ($F(2,19) = 0.018$, $p = 0.98$), indicating that neither the type of guidance (human or autonomous) nor its perceived source affected objective trust.


max 84
80
70
60
50
40
30
20
10
min 7
Self-reported Trust
tH
tArH
tArA
Experimental Group
tH
tArH
tArA

291

Figure 4 Boxplot of self-reported trust across experimental groups. Control group omitted as these subjects did not use a navigational guidance system. Minimum and maximum possible scores from Jian et al.'s (2000) trust questionnaire are shown.

Figure 4 shows self-reported trust, collected in the post-test questionnaire (Jian et al., 2000) for each subject (except Control subjects). Again, we applied a one-way ANOVA and found no significant differences across groups ($F(2,19) = 0.082$, $p = 0.92$), further suggesting similar trust between human and autonomous guidance (perceived or actual).

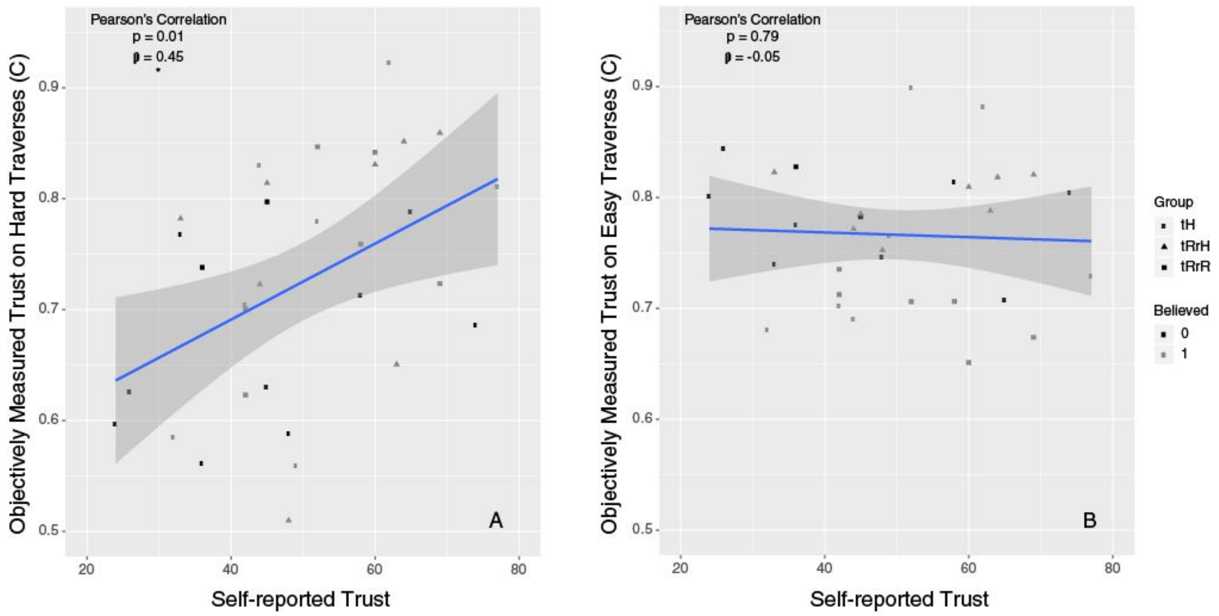


Figure 5 Correlations between self-reported trust and objective measured trust. Panel A shows data from the 'hard' traverses and panel B for 'easy' traverses (see Table 4). The correlation coefficients (ρ), line-of-best-fit, and the associated confidence band are shown. All subjects that completed the post-test self-reported trust questionnaire (i.e., even that did not believe the source of guidance they were told, but not control subjects, n=32) were included as we are examining simply if self-reported trust was associated with objective trust. However, whether the subject believed the guidance or did not is depicted with the color of the marker. Shapes show the three experimental groups (tH, tRrH, tRrR).

We considered the relationship between objective and self-reported trust separated based upon the traverse difficulty (easy or hard). A significant positive correlation was only found between self-reported trust and average objective trust on 'hard' traverses, shown in Figure 5A ($p = 0.01 < \alpha =$

0.025 after Bonferroni correction, ρ(30) = 0.45). On the 'easy' traverses (Figure 5B), the correlation between objective and subject trust was not significant (p = 0.79, ρ(30) = -0.05). This suggests that self-reported trust correlates with objective trust only when the human operator is more likely to require help from the system (i.e., on the 'hard' traverses).

Discussion

Our results indicate that substantial intervention may not be necessary to have humans trust an autonomous guidance system a similar amount to that which they trust a human one. This is consistent with the Ma and Kaber study (2007), which found that, after use, their subject pool had similar trust in a human vs autonomous guidance system when competency of the systems was equal. Here, we control for competency of the guidance system by deceiving some subject groups. For example, the 'told autonomous, received autonomous' group [tArA] had very similar trust that for the 'told autonomous, received human' group [tArH]. Our findings of consistent trust between human and autonomous guidance are an interesting contrast to the several studies showing that conveying the same information in a more humanistic manner results in higher trust (Li & Yeh, 2010; Nass & Moon, 2000; Stedmon et al., 2007). Here, the presentation of the guidance was identical. Instead, the source of the guidance being human or autonomous was conveyed in instructions by the experimenter. It may be that the impact of humanistic interactions on trust only applies to how the information is presented and not on its perceived origin.

Considering our objectively measured trust metric to quantify how compliant subjects were with the system, our finding of a significant correlation between objective trust and self-reported trust is consistent with Muir and Moray's (1996) finding that trust correlated with automation use. Specifically, when subjects report a higher level of trust in the system, they are more likely to use or comply with the guidance provided. The results are also consistent with the trending positive relationship between trust and system use found by Dzindolet (2003). Our findings caveat this relationship in that self-reported trust only correlated when the task was 'hard', but not when it was an 'easy' traverse. This suggests that operator trust is important for modulating behavior when the task is challenging and the operator is more pressed to decide whether to comply with the system's guidance. Furthermore, it appears this relationship is not dependent upon having autonomous versus human navigational assistance, but instead is a property of integrating guidance when performing difficult tasks.

Additionally, dynamic changes in trust over time can more easily be assessed via the methods we propose. In the Lewandowsky (2000) and Yang (2017) studies, the subject is polled for trust multiple times throughout the experiment. In contrast, our objective measure of trust can be used in real time in an unobtrusive manner, without needing to poll subjects at discrete intervals. Our finding that each subject's average objective trust measure correlated with their post-experiment questionnaire self-reported measure of trust helps validate this unobtrusive measure. While the caveat that the correlation only exists for hard traverses could be considered a limitation of the unobtrusive, objective measure, understanding trust during difficult tasks may be most critical. We note that this unobtrusive, objective measure of trust does have some limitations. It does not capture all aspects of trust, but instead focuses specifically on the compliance aspect. While continuous, at any instant, the measure may be misleading. For example, right before reaching a goal the subject is likely to drive aligned with the guidance straight towards the goal, even if their trust in the guidance is very low.

A key limitation of our study is not having sufficiently many subjects in either 'told human' groups. We intended to compare the control group as compared to human guidance [CcH] to the original 'told human, received human' [tHrH] group. This would have enabled us to investigate how providing guidance that was perceived as being from the human would impact trust behaviors. Unexpectedly, in the post-test questionnaire many of the subjects in the 'told human' groups reported not believing the experimenter (even when truthful). Consequently, our results may be influenced by demand characteristics, where subject behaviors may have been consciously or unconsciously influenced by their beliefs regarding the intent of the experiment and the type of guidance they received. We intentionally did not have subjects interact with, or even see, the guidance human research assistant because we did not want the subject's responses to be influenced by any immediate judgments of character. Our intention was to capture responses when interacting with another, unseen human (i.e., analogous to an astronaut with a ground controller). However, allowing for a brief interaction may have helped subjects in the 'told human' groups to believe the experimenter. By pooling all six subjects that were told human and believed it [H], we were still able to compare outcomes to the groups in which subjects were told the guidance was autonomous. We found the source they were told (autonomous versus human) did not have a significant impact upon objective or self-reported trust metrics.

We are also limited by the diversity of our subject pool. Previous studies have shown age to significantly impact trust in automation with older subjects reporting higher trust in automation (Donmez et al., 2006; Ho et al., 2005; Kircher & Thorslund, 2009). Our subject pool was young and not representative of the population in general. Nonetheless, we observe substantial variation between individuals in both objective trust (Figure 3) and self-reported trust (Figure 4, where in both cases boxplots were used to visualize these inter-individual distributions). This inter-individual variation may be attributed to innate time-invariant aspects of trust (Schaefer et al., 2016) (though such differences are not well studied (Hancock et al., 2009)), that are independent of operator age. Further study would be prudent to confirm that our results extend to the population at large.

Contributions

We have implemented a method for continuously, unobtrusively and objectively measuring trust in an autonomous guidance system. While the presence of the autonomous guidance system changes subject behavior in a statistically significant manner, the origin of the guidance does not appear to matter: Our subject pool was agnostic to guidance source (true or perceived).

There was a positive correlation between self-reported trust and objectively measured trust on difficult traverses, implying that our users' conscious trust in the system was consistent with how they complied with the system. Lack of correlation between self-reported trust and objectively measured trust on easy traverses suggests that although self-reported trust can be a predictor of compliance, when the task can be done easily without the system, the association between trust and compliance fades.

Key Points

- When we control for the performance of the system, human operators do not show significant differences in trust (self-reported or objectively measured) between human or autonomous guidance systems.
- Our method for objectively quantifying trust is unobtrusive and can be passively measured. This metric of trust correlates significantly with self-reported trust (assessed with a standard questionnaire) on difficult tasks. This suggests that, particularly for challenging tasks, dynamic changes in trust can be unobtrusively quantified with our objective metric.

References

Bindewald, J. M., Rusnock, C. F., & Miller, M. E. (2018). Measuring Human Trust Behavior in Human-Machine Teams. In D. N. Cassenti (Ed.), *Advances in Human Factors in Simulation and Modeling* (Vol. 591, pp. 47–58). Springer International Publishing. https://doi.org/10.1007/978-3-319-60591-3_5

Chiou, E. K., & Lee, J. D. (2015). Beyond Reliance and Compliance: Human-Automation Coordination and Cooperation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 195–199. https://doi.org/10.1177/1541931215591040

Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., & Yanco, H. (2013). Impact of robot failures and feedback on real-time trust. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 251–258. https://doi.org/10.1109/HRI.2013.6483596

Dixon, S. R., & Wickens, C. D. (2006). Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(3), 474–486. https://doi.org/10.1518/001872006778606822

Donmez, B., Boyle, L. N., Lee, J. D., & McGehee, D. V. (2006). Drivers’ attitudes toward imperfect distraction mitigation strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *9*(6), 387–398. https://doi.org/10.1016/j.trf.2006.02.001

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(1), 79–94. https://doi.org/10.1518/0018720024494856

Hancock, P.A., Hancock, G. M., & Warm, J. S. (2009). Individuation: The *N* = 1 revolution. *Theoretical Issues in Ergonomics Science*, *10*(5), 481–488. https://doi.org/10.1080/14639220903106387

Hancock, Peter A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Ho, G., Wheatley, D., & Scialfa, C. T. (2005). Age differences in trust and reliance of a medication management system. *Interacting with Computers*, *17*(6), 690–710. https://doi.org/10.1016/j.intcom.2005.09.007

Holden, K., Ezer, N., & Vos, G. (2013). *Evidence report: Risk of inadequate human-computer interaction*. NASA Technical Report Server. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20130014064.pdf

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Kircher, K., & Thorslund, B. (2009). Effects of road surface appearance and low friction warning systems on driver behaviour and confidence in the warning system. *Ergonomics*, *52*(2), 165–176. https://doi.org/10.1080/00140130802277547

Kochenderfer, M. J. (2015). *Decision making under uncertainty: Theory and application*. The MIT Press.

Lacson, F. C., Wiegmann, D. A., & Madhavan, P. (2005). Effects of Attribute and Goal Framing on Automation Reliance and Compliance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(3), 482–486. https://doi.org/10.1177/154193120504900357

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104–123. https://doi.org/10.1037/1076-898X.6.2.104

Li, Y.-M., & Yeh, Y.-S. (2010). Increasing trust in mobile commerce through design aesthetics. *Computers in Human Behavior*, *26*(4), 673–684. https://doi.org/10.1016/j.chb.2010.01.004

Ma, R., & Kaber, D. B. (2007). Effects of in-vehicle navigation assistance and performance on driver trust and vehicle control. *International Journal of Industrial Ergonomics*, *37*(8), 665–673. https://doi.org/10.1016/j.ergon.2007.04.005

McGuire, S., Walker, M., McGinley, J., Ahmed, N., Szafir, D., & Clark, T. (2018). *TRAADRE: TRust in Autonomous ADvisors for Robotic Exploration*. 5.

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(3), 520–534. https://doi.org/10.1177/0018720812465081

Merritt, S. M., & Ilgen, D. R. (2008). Not All Trust Is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(2), 194–210. https://doi.org/10.1518/001872008X288574

Merritt, S. M., Lee, D., Unnerstall, J. L., & Huber, K. (2015). Are Well-Calibrated Users Effective Users? Associations Between Calibration of Trust and Performance on an Automation-Aided Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(1), 34–47. https://doi.org/10.1177/0018720814561675

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460. https://doi.org/10.1080/00140139608964474

NASA. (2019). *We do not know how to quantify overall human-automation-robotic system performance to inform and evaluate system designs to ensure safe and efficient space mission operations.* NASA Human Research Roadmap. https://humanresearchroadmap.nasa.gov/Gaps/gap.aspx?i=325

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Sheridan, T. B. (2002). *Humans and automation: System design and research issues*. Wiley [u.a.].

Stedmon, A. W., Sharples, S., Littlewood, R., Cox, G., Patel, H., & Wilson, J. R. (2007). Datalink in air traffic management: Human factors issues in communications. *Applied Ergonomics*, *38*(4), 473–480. https://doi.org/10.1016/j.apergo.2007.01.013

Voros, J. L., McGinley, J., McGuire, S., Walker, M. E., Karki, P., Ahmed, N., Szafir, D., & Clark, T. K. (2020, March 7). *Trust in an Autonomous Guidance System for a Planetary Rover Task*. IEEE Aerospace Conference, Big Sky, MT.

Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017). Evaluating Effects of User Experience and System Transparency on Trust in Automation. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI ’17*, 408–416. https://doi.org/10.1145/2909824.3020230

Biographies

Jamie L Voros

Current affiliation: Ann & H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder

Highest degree obtained: M. S. Aerospace Engineering Sciences, 2018, University of Colorado Boulder

Jamison McGinley

Current affiliation: Ann & H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder

Highest degree obtained: B. S. Aerospace Engineering Sciences, 2020, University of Colorado, Boulder.

Steve McGuire

Current affiliation: Department of Electrical and Computer Engineering, University of California Santa Cruz

Highest degree obtained: Ph. D., Ann & H.J. Smead Department of Aerospace Engineering Sciences, 2019, University of Colorado Boulder

Michael E Walker

Current affiliation: Department of Computer Science, ATLAS Institute, University of Colorado Boulder

Highest degree obtained: M. S., Computer Science, 2018, The University of Colorado Boulder

Priyanka Karki

Current affiliation: Ann & H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder

Highest degree obtained: Candidate for B. S. Computer Science, 2021, University of Colorado Boulder

Nisar Ahmed

Current affiliation: Ann & H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder

Highest degree obtained: Ph. D Mechanical Engineering, 2012, Cornell University.

Daniel Szafir

Current affiliation: Department of Computer Science, ATLAS Institute, University of Colorado Boulder

Highest degree obtained: Ph. D. Computer Science, 2015, University of Wisconsin-Madison.

Torin K Clark

Current affiliation: Ann & H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder

Highest degree obtained: Ph. D. Humans in Aerospace, 2013, MIT.