Convergence Across Behavioral and Self-report Measures Evaluating Individuals' Trust in an Autonomous Golf Cart

Jenna E. Cotter, Emily H. O'Hear, R. Cooper Smitherman, Addison B. Bright, and Nathan L. Tenhundfeld
Department of Psychology,
The University of Alabama in Huntsville, Huntsville, USA
jc0277@uah.edu, eho0002@uah.edu, rcs0024@uah.edu, abb0040@uah.edu, nlt0006@uah.edu

Jason Forsyth, Nathan R. Sprague, and Samy El-Tawab
College of Integrated Science and Engineering
James Madison University, Harrisonburg, USA
forsy2jb@jmu.edu, spragunr@jmu.edu, eltawass@jmu.edu

***Abstract*— As automation is becoming more prevalent across everything from military and healthcare settings to everyday household items, it is necessary to understand the nature of human interactions with these systems. One critically important element of these interactions is user trust, as it can predict automated systems' safe and effective use. Past research has evaluated individuals' trust in automation through a host of different assessment techniques such as self-report, physiological, and behavioral measures. However, to date, there has been little evaluation of the convergence across these measures in a real-world environment. Convergence across measures is a useful tool in understanding the mechanisms by which a cognitive construct is impacted and providing greater confidence that any single measure is evaluating what it purports to measure. The present study used an autonomous golf cart that drove participants to different locations around the campus of James Madison University while a camera recorded them. In addition, participants were given the AICP-R and TOAST to evaluate their complacency potential and trust, respectively. Researchers coded videos for verification/checking behaviors (i.e., participants looked at or interacted with the GUI used to control the cart) and nervous behaviors (i.e., bracing, fidgeting, etc.). Additionally, environmental 'obstacles' such as pedestrians, food-delivery robots, and construction were also coded for by watching a front-facing camera. Results indicate a disconnect between the self-report and behavioral measures evaluating trust. However, there was a relationship between the coded nervous behaviors and verification behaviors and a relationship between those and the presence of obstacles. This lack of convergence across measures indicates a need for future research to understand whether this non-convergence represents shortcomings with the measures themselves, the existing definition of trust as a construct, or perhaps indicates that there is a nuance that can be afforded by some measures over another.**

Keywords—Automation, Trust, Convergence, Complacency, Verification, Autonomous Vehicles

I. INTRODUCTION

Given recent advances and the rise in self-driving vehicles on roadways, it is essential to understand how individuals interact with these systems. Automation, in general, is present in everyday life and is used in settings such as healthcare, the military, aviation, and household items. As much as automation can reduce workload, increase productivity, and reduce costs, the increasing complexity of automation also presents unique challenges. The recent development of self-driving vehicles came with a cost [1]. In a fatal accident in 2016, a Tesla in "Autopilot" mode failed to apply the brakes when a tractor-trailer made a left turn in front of the Tesla [2]. In 2019, a Tesla going 60 miles per hour slammed into a parked firetruck [3]. In these instances, it is reasonable to attribute user overtrust in the system to their failure to intervene appropriately before a disaster.

Trust is a useful measure in that it serves as a predictor of a system's use, misuse, and disuse [4]. Trust is defined as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [5]. The continued investigation of trust in human-automation interactions is critical for promoting safe interactions and enhancing human-machine team performance. When designing systems, trust must be *calibrated* rather than elevated. A user's trust in a system must correspond with the actual capabilities of that system. Overtrust refers to when the user's trust in the system is higher than the machine's actual capabilities [6][7]. Users who overtrust a system can be conditioned to become complacent and misuse the system, possibly leading to the loss of pricey equipment or human lives [4], [8].

Complacency refers to monitoring automation with less vigilance or less frequently because of a greater level of trust in the system [9]. Oftentimes, operators who become complacent are influenced by automation biases [10] [11]. Users that expect automation to perform without error tend to place too much trust in the machine [12]. To offset overtrust and appropriately calibrate trust, trust dampening approaches such as providing timely warnings, requesting assistance, or conveying system limitations can be implemented to lower expectations when too much faith has been placed into the system [4] [13].

On the other hand, undertrust refers to when the user's perception of the system's trustworthiness is lower than the

978-1-6654-5111-6/22/$31.00 ©2022 IEEE

machine's actual capabilities [4]. Undertrust can lead to the disuse of a system [8], [14], [15]. In situations of undertrust, trust can be calibrated through trust repair strategies such as apologizing, denial, blaming, or gas-lighting [4], [16] [17]. When designing a system, keeping these approaches in mind will help calibrate trust within human-automation interactions.

Ultimately, trust calibration is necessary for appropriate reliance on a system. However, to calibrate trust, there must be an accurate understanding of user trust. Previous research has evaluated individuals' trust in automation through different assessment techniques such as subjective self-reports [18], [19], physiological [20], and behavioral measures [1], [15].

The remainder of the paper is organized as follows: Section II discusses the current state of the art of the autonomous golf cart used for our research study. Section III explains current measurement techniques used on autonomous systems. In section IV, we discuss the methods used in our research, including the number of participants, materials, procedures, and coding procedure. We follow with section V, explaining the results. Finally, we conclude the paper by discussing our research results and highlighting future work directions in section VI.

II. Overview of Current Autonomous Cart

Our study uses an autonomous golf-cart prototype created by James Madison University Autonomous Golf Cart (JACart) team. As shown in Figure 1, the autonomously driven golf cart has several components: a 3D LiDAR, Zed cameras, hardware controller, UI control, and a laptop running Robot Operating System (ROS). This work is part of an effort to develop a prototype taxi system as a testbed for experimenting with features to improve accessibility and comfortability.

FIGURE 1. Autonomous Vehicle cart used

III. Current Measurement Techniques

A. Self-Reports

Self-reports are inherently subjective and easy to administer. These assessment techniques typically involve asking individuals a series of questions that relate to their perception of the system with which they were interacting. One of the most commonly used trust measures is the Trust in Automation Scale, which evaluates respondents' impression/perception of the system [18]. While this scale is widely used by the research community (cited over 1200 times at the writing of this manuscript), evidence of positive biases producing ceiling effects [21] and insufficient reliability [18] raise concerns about its use. However, other scales have been developed which focus on minimizing administration biases while also reporting stronger psychometric properties making them a key candidate for use in individual differences studies (e.g., [22]). While self-report measures are easy to administer, they rely on the individual's responses which may be faulty or inaccurate as they are subjective and may be impacted by any number of biases due to the scale construction or experimental demands [23]–[25]. There has been an interest in establishing behavioral indicators of trust.

B. Behavioral

Behavioral measures are collected during an individual’s interactions with a system. By collecting data during the interaction, trust assessments may be more accurate than subjective assessments that require a reflection after the interaction. Additionally, this behavioral approach to trust assessment allows for assessment in real time. This real-time understanding of a user’s trust in a system can allow for a more rapid response from the system itself to help recalibrate the user’s trust. The use of behavioral measures is non-invasive and can therefore be used in a much more passive monitoring environment. Previous research has used behaviors such as verification [26], [27], intervention [15], [28], and secondary task engagement [29] as indicators of one’s trust.

C. Convergence

Convergence is the agreement among measures that assess a given construct. For instance, an experiment may use both subjective self-report and behavioral measures of trust to establish the level of agreement across those measures. One would anticipate that measures that assess the same construct should yield converging results, such that both the subjective self-report and behavioral data should show similar patterns of results. This approach of converging measures has historically been used to assess other cognitive constructs and evaluate the validity of a given measure [30]. When convergence is not evident across measures, this potentially indicates that the construct is being measured inadequately, or perhaps the construct's definition fails to account for nuance in the construct. The complex nature of measuring trust means that it is ripe for exploration into the degree of convergence across measures [16], [30].

IV. Methods

Participants. Participants ($N = 26$) were recruited from an available participant pool at James Madison University (JMU), Harrisonburg, Virginia, USA.

Materials. This study used an EZ-Go Golf Cart, which was adapted to operate the carts' brakes, accelerator, and steering autonomously [31], [32]. A Robot Operation System (ROS) program filters information of 3-D LiDAR (Velodyne Puck) to detect obstacles, defined routes, and the cart's location. There is a graphical user interface (GUI) located inside the cart so that passengers can select destinations and potentially void operations if necessary. The GUI also provides a visual representation of the path from start to end. The Automation Induced Complacency Potential Rating (AICP-R) is used to evaluate the complacency potential of their interactions with automation. The AICP-R consists of two factors: alleviating workload and monitoring [6]. The Trust of Automated Systems Test (TOAST) assesses the individuals' overall trust in understanding the system and the performance of automation [22].

Procedures. Before beginning the experiment, participants are asked to read the informed, voluntary consent form and sign if they agree to participate. Participants were informed that this study aims to understand better passengers' feelings of trust and reliability in autonomous vehicles. After participants provided consent, they were encouraged to ask any questions about the experiment. Participants were then given a tablet instructing them to complete the AICP-R. After the pre-ride survey was conducted, the experimenter fitted the participants with a pulse-oximeter; however, this physiological data was collected for another project and thus will not be analyzed here.

Participants were shown the cart and asked to sit in the passenger seat of the golf cart while the researcher informed them of the GUI interface. They were informed about the different destinations (i.e., gym, café, clinic, mall, movie theater) or obstacles (i.e., food robots, pedestrians, cars, etc.) that they may encounter on campus during the ride. Prior to the ride, the researcher reminded participants that they should pretend that the experimenter was not there and would not generally intervene unless there was a safety issue. Following this explanation, participants were asked to begin a course from home to mall and then from mall to home. Once the cart returned home, participants were given the opportunity to select a different location for an additional ride.

After the ride, a semi-structured interview was completed while the participant was in the cart, which allowed their responses to be recorded and later transcribed. The interview consisted of questions about the ride (i.e., stress, reliability, familiarity with autonomous vehicles, etc.) such as "Did you feel comfortable and/or at ease during the ride? If so, what made you feel that way?" Lastly, participants were given a tablet to complete the post-ride survey (TOAST). After the questionnaire was completed, the pulse oximeter was removed, and participants were thanked for their participation in the study.

Coding Procedure. Researchers coded the videos independently by evaluating participants' actions and recording timestamps from the listed categories (i.e., obstacles, GUI interactions, and participants' behaviors) (Table 1). An event was considered to have occurred when at least two of the three coders recorded the event within 2 seconds of each other, and at least 15 seconds since the last time there was a recorded event of that type. This was important to guarantee agreement among researchers while also preventing over-coding similar events that were recorded by researchers.

The behaviors selected as indications of our representation of trust or distrust within this study align with previous research, which defines these actions as displacement behaviors included within the Ethological Coding System for Interviewers (ECSI) [33]. These displacement behaviors signify feelings of stress and anxiety, which may betray what individuals report in self-report measures, coined by the term emotional leakage [33], [34]. For example, research studies have shown that displacement behaviors result from individuals experiencing anxiety or stress [35]–[37]. Behaviors such as touching one's face or fidgeting with hands fall within the ECSI and may further serve to demonstrate anxiety that may not be reported by the individual, which could be indicative of his/her trust.

TABLE I. OBSTACLES, BEHAVIORS, AND INTERACTIONS

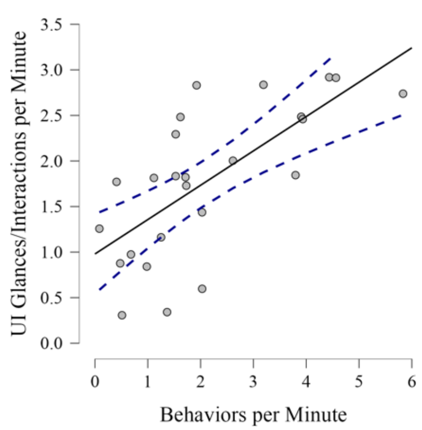
Coding Categories			
Obstacles	***GUI Interaction***	***Participant Behaviors***	***Participant Behaviors Cont.***
Food Robots	Looking at GUI	Bracing	Laughing
Pedestrians	Interacting with GUI	Glancing	Interaction with phone
Cars		Darting eyes	Noticeable gasp
Curves in the road		Noticeable exhale	Excitement
Construction		Joking around	Interacting with pedestrians
		Talking to self	Fidgeting
		Slouching	Crossing legs
		Nodding	Touching face
		Pointing	Leaning forward
		Smiling	Messing with hands
		Mouth movement	Hair flip

V. Results

The statistical software JASP [38] was utilized to analyze the data to report frequentist statistics and Bayes Factors (BF_{10}). A BF_{10} between 1-3 is classified as anecdotal, 3-10 is moderate, 10-30 is strong, and > 30 is very strong evidence [39]. The advantage of using Bayesian statistics over frequentists is the ability to gain evidence favoring the null hypothesis and discriminate between the "absence of evidence" (i.e., inconclusive, or insignificant) or "evidence of absence" (i.e., support H0) [40]. Additionally, Bayesian analyses are more immune to issues of small sample sizes [39]

Individuals had an average AICP-R score of 3.552 (*SD* = 0.336) (out of 5), with an average of 4.024 (*SD* = 0.543) for alleviating workload and an average of 3.080 (*SD* = 0.370) for monitoring subscale. The average reported trust in the system on TOAST was 5.893 (*SD* = 0.466). The mean number of obstacles coded for every participant was 17.080 (*SD* = 7.810), number of behaviors 23.240 (*SD* = 14.475), and glances/interactions with UI average to 17.760 (*SD* = 7.865). There was no correlation found between the AICP-R and TOAST scores, $r(25) = .206$, $p = .324$, $BF_{10} = .393$. For all DV counts (e.g. number of obstacles encountered), a standardized value was calculated of number per minute, in order to control for length of time. There was not a significant correlation between the number of obstacles that individuals encountered and their reported trust, collected with TOAST, $r(25) = -0.139$, $p = .508$, $BF_{10} = .306$. No correlation was found between TOAST and the number of behaviors coded, $r(25) = .046$, $p = .827$, $BF_{10} = .254$, or between the TOAST and the number of glances with the UI, $r(25) = .117$, $p = .579$, $BF_{10} = .287$. Indicating moderate evidence in favor of the null hypothesis for the correlation between TOAST and the number of behaviors, in addition to the number of glances with UI.

FIGURE 2. Correlation Between Behaviors and UI Glances/Interactions



Note: The dashed lines represent the 95% confidence interval for the regression line fit onto the data.

There was a significant correlation between the number of obstacles and number of behaviors coded, $r(25) = .500$, $p = .011$, $BF_{10} = 5.280$, indicating moderate evidence. A significant correlation was found between the number of obstacles coded and the number of glances with UI, $r(25) = .618$, $p < .001$, $BF_{10} = 41.788$, with very strong evidence in favor of the alternative hypothesis. There was also a significant correlation between the number of behaviors and glances with the UI, $r(25) = .695$, $p < .001$, $BF_{10} = 272.491$ (Figure 1), with extreme evidence in favor of the alternative hypothesis. Finally, there was no correlation between the monitoring subscale, and number of glances/interactions with UI $r(25) = -.017$, $p = .938$, $BF_{10} = .249$.

VI. Discussion

The study's goal was to examine further individuals' trust in automation and the relationships between different measures used to assess trust. Trust is a construct that can be measured through numerous assessment techniques, and this study utilized self-reporting measures and behavioral measures [4], [16]. The results from the present study show no correlation between the number of coded behaviors and TOAST. Our results indicate moderate evidence in favor of the null hypothesis between the TOAST score and the number of coded behaviors. This lack of convergence could be explained in several ways. First, it may simply be a function of the fact that our participants were highly trusting of the system. The individuals in our study had an average score of 5.893 out of 7 on the TOAST. As such, it is likely the case that any behaviors which could be considered 'anxious' behaviors in response to low trust, could simply have been motivated by some other antecedent.

Alternatively, it is possible that the behaviors and self-report were assessing different constructs altogether and that behaviors, despite being used in the past, may not be a great indicator of trust. One reason is that behaviors obviously have many motivators beyond simply trust in an automated system. One final potential explanation for these results is the timing of them. Perhaps participants did not trust the system at the moment; however, following the trial, participants were able to reflect upon their experience, realize there were no adverse events, and thus indicated a higher level of trust. Unfortunately, our results cannot speak directly to any of these potential explanations.

Our results indicate that scores of individuals on the AICP-R monitoring subscale were surprisingly not correlated with the number of glances with GUI, which we believed to be a good measure of actual monitoring. The lack of correlation between AICP-R and glances at the GUI potentially indicates a disconnect between one's level of possible complacency toward a system and the level of trust displayed toward that same system. Additionally, the GUI in this study may not have provided enough information to the individual to warrant them monitoring it at all, thus leading to nonsignificant results. However, there was only anecdotal support for the null hypothesis, and we would therefore caution over-extrapolation of the results.

Despite the aforementioned lack of convergence, three significant results regarding behaviors exhibited by the individuals were found:

1. A significant correlation between the number of obstacles and coded behaviors was found. This is expected in that obstacles encountered would likely be a period of stress given the anticipation and uncertainty surrounding how the vehicle would handle the situation.
2. There was a significant correlation between the number of obstacles and the number of times a participant glanced at the GUI. Similarly, this is to be expected given that the more frequently one would interact with a system, the more often they may perform checking/verification behaviors.
3. There was a significant correlation with extreme evidence in favor of the alternative hypothesis, between the number of behaviors exhibited by the individuals and the number of glances at/interactions with the GUI, which shows a degree of convergence across these behavioral measures.

The displacement behaviors selected in this study were correlated with the external obstacles and the number of glances at the GUI, thus further demonstrating that the trust was misplaced as a function of the external attributes rather than trust in the system itself.

Evaluating convergence across measures provides a more complete picture of a given construct. The results demonstrated a lack of convergence across the AICP-R, TOAST, and behavioral measures which could suggest that our definition of trust may be insufficiently nuanced to account for discrepant results found across self-report and behavioral measures. However, because of the high degree of convergence across the behavioral measures, it appears as though our behavioral measures are assessing a set construct. Whether this construct is in fact trust warrants further exploration. Although research has shown a correlation between displacement behaviors and stress [33], [34], [43] there was no indication that this stress is indicative of trust as defined by the TOAST self-report. This could imply that the measures are measuring two different aspects of trust or constructs entirely.

This study is somewhat limited in the assessment techniques used, and in the environment in which they were tested. These results do not signify that current assessment techniques are invalid, nor that there would not be convergence across measures in other environments. However, the results of this study necessitate further investigation into why a lack of convergence was observed. Future research should aim to determine whether the behavioral measures used accurately measure trust. Should these results be replicated in other environments, and with other assessment techniques, the definition of trust as a construct might require reevaluation.

References

[1] N. L. Tenhundfeld, E. J. de Visser, A. J. Ries, V. S. Finomore, and C. C. Tossell, “Trust and Distrust of Automated Parking in a Tesla Model X,” *Hum. Factors*, vol. 62, no. 2, pp. 194–210, 2020, doi: 10.1177/0018720819865412.

[2] R. Abrams and A. Kurtz, “Joshua Brown, Who Died in Self-Driving Accident, Tested Limits of His Tesla,” *The New York Times*, 2016.

[3] N. Boudette and N. Chokshi, “Tesla Autopilot Faces U.S. Inquiry After Series of Crashes - The New York Times.” .

[4] E. J. de Visser *et al.*, “Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams,” *Int. J. Soc. Robot.*, vol. 12, no. 2, pp. 459–478, 2020, doi: 10.1007/s12369-019-00596-x.

[5] J. D. Lee and K. A. See, “Trust in Automation: Designing for Appropriate Reliance,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 46, no. 1, pp. 50–80, 2004, doi: 10.1518/hfes.46.1.50_30392.

[6] S. M. Merritt *et al.*, “Automation-induced complacency potential: Development and validation of a new scale,” *Front. Psychol.*, vol. 10, no. 225, pp. 1–13, 2019, doi: 10.3389/fpsyg.2019.00225.

[7] S. M. Merritt, J. L. Unnerstall, D. Lee, and K. Huber, “Measuring Individual Differences in the Perfect Automation Schema,” *Hum. Factors*, vol. 57, no. 5, pp. 740–753, 2015, doi: 10.1177/0018720815581247.

[8] R. Parasuraman and V. Riley, “Humans and Automation: Use, Misuse, Disuse, Abuse,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 39, no. 2, pp. 230–253, 1997, doi: 10.1518/001872097778543886.

[9] R. Parasuraman, R. Molloy, and I. L. Singh, “Performance Consequences of Automation-Induced ‘Complacency,’” *Int. J. Aviat. Psychol.*, vol. 3, no. 1, pp. 1–23, 1993, doi: 10.1207/s15327108ijap0301_1.

[10] J. E. E. Bahner, A. D. A. D. Hüper, and D. H. Manzey, “Misuse of automated decision aids: Complacency, automation bias and the impact of training experience,” *Int. J. Hum. Comput. Stud.*, vol. 66, no. 9, pp. 688–699, 2008, doi: 10.1016/j.ijhcs.2008.06.001.

[11] J. E. Bahner, M. F. Elepfandt, and D. Manzey, “Misuse of diagnostic aids in process control: The effects of automation misses on complacency and automation bias,” *Proc. Hum. Factors Ergon. Soc.*, vol. 2, pp. 1330–1334, 2008, doi: 10.1177/154193120805201906.

[12] J. B. Lyons and S. Y. Guznov, “Individual differences in human–machine trust: A multi-study look at the perfect automation schema,” *Theor. Issues Ergon. Sci.*, vol. 20, no. 4, pp. 440–458, Jul. 2019, doi: 10.1080/1463922X.2018.1491071.

[13] R. B. Jackson and T. Williams, “Language-Capable Robots may Inadvertently Weaken Human Moral Norms,” in *Proceedings of the Companion of the 14th ACM/IEEE International Conference on Human-Robot Interaction (alt. HRI)*, 2019, pp. 401–410, doi: 10.1109/HRI.2019.8673123.

[14] K. A. Hoff and M. Bashir, “Trust in Automation : Integrating Empirical Evidence on Factors That Influence Trust,” *Hum. Factors*, vol. 57, no. 3, pp. 407–434, 2015, doi: 10.1177/0018720814547570.

[15] N. L. . Tenhundfeld, E. J. de Visser, K. S. . Haring, A. J. Ries, V. S. Finomore, and C. C. Tossell, “Calibrating trust in automation through familiarity with the autoparking feature of a Tesla Model X,” *J. Cogn. Eng. Decis. Mak.*, vol. 13, no. 4, pp. 279–294, 2019, doi: 10.1177/1555343419869083.

[16] S. C. Kohn, E. J. de Visser, E. Wiese, Y.-C. Lee, and T. H. Shaw, “Measurement of Trust in Automation: A Narrative Review & Reference Guide,” *Front. Psychol.*, no. October, p. 85, 2021, doi: 10.3389/fpsyg.2021.604977.

[17] M. Nayyar and A. R. Wagner, “When Should a Robot Apologize? Understanding How Timing Affects Human-Robot Trust Repair,” in *International Conference on Social Robotics*, 2018, pp. 265–274.

[18] J.-Y. Jian, A. M. Bisantz, C. G. Drury, and J. Llinas, “Foundations for an Empirically Determined Scale of Trust in Automated Systems,” *Int. J. Cogn. Ergon.*, vol. 4, no. 1, pp. 53–71, 2000.
[19] M. T. Dzindolet, S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck, “The role of trust in automation reliance,” *Int. J. Hum. Comput. Stud.*, vol. 58, no. 6, pp. 697–718, 2003, doi: 10.1016/S1071-5819(03)00038-7.
[20] H. M. Khalid *et al.*, “Exploring psycho-physiological correlates to trust: Implications for human-robot-human interaction,” in *Proceedings of the Human Factors and Ergonomics Society*, 2016, pp. 696–700, doi: 10.1177/1541931213601160.
[21] R. S. Gutzwiller, E. K. Chiou, S. D. Craig, C. M. Lewis, G. J. Lematta, and C.-P. Hsiung, “Positive bias in the ‘Trust in Automated Systems Survey’? An examination of the Jian et al. (2000) scale,” 2019, doi: 10.1177/1071181319631201.
[22] H. Wojton, S. Lane, and D. Porter, “Initial Validation of the Trust of Automated Systems Test (TOAST),” *J. Soc. Psychol.*, pp. 1–16, 2020.
[23] R. S. Gutzwiller, E. K. Chiou, S. D. Craig, C. M. Lewis, G. J. Lematta, and C. Hsiung, “Positive bias in the ‘Trust in Automated Systems Survey’? An examination of the Jian et al. (2000) scale,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 63, no. 1, pp. 217–221, 2019, doi: 10.1177/1071181319631201.
[24] E. C. Poulton, “Models for Biases in Judging Sensory Magnitude,” *Psychol. Bull.*, vol. 86, no. 4, pp. 777–803, 1979.
[25] S. S. Stevens, “On the psychophysical law,” *Psychol. Rev.*, vol. 64, no. 3, pp. 153–181, 1957, doi: 10.1037/h0046162.
[26] M. L. Bolton, E. J. Bass, and R. I. Siminiceanu, “Using formal verification to evaluate human-automation interaction: A review,” *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, vol. 43, no. 3, 2013, doi: 10.1109/TSMCA.2012.2210406.
[27] N. Ezer, A. D. Fisk, and W. A. Rogers, “Reliance on Automation as a Function of Expectation of Reliability, Cost of Verification, and Age,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 51, no. 1, pp. 6–10, 2007, doi: 10.1177/154193120705100102.
[28] K. Tomzcak *et al.*, “Let Tesla Park Your Tesla: Driver Trust in a Semi-Automated Car,” 2019, doi: 10.1109/SIEDS.2019.8735647.
[29] V. A. Banks, A. Eriksson, J. O’Donoghue, and N. A. Stanton, “Is partially automated driving a bad idea? Observations from an on-road study.,” *Appl. Ergon.*, vol. 68, pp. 138–145, 2018.
[30] N. L. Tenhundfeld and J. K. Witt, “Distances on hills look farther than distances on flat ground: Evidence from converging measures,” *Attention, Perception, Psychophys.*, vol. 79, no. 4, pp. 1165–1181, 2017, doi: 10.3758/s13414-017-1305-x.
[31] S. S. El-Tawab, N. Sprague, and A. Mufti, “Autonomous Vehicles: Building a Test-bed Prototype at a Controlled Environment,” *IEEE World Forum on Internet of Things, WF-IoT 2020 - Symposium Proceedings*. pp. 1–6, 2020, doi: 10.1109/WF-IoT48130.2020.9221222.
[32] S. S. El-Tawab, N. Sprague, M. Stewart, M. Pareek, and P. Zubov, “Enhanced Interface for Autonomously driven Golf Cart in a Networked Controlled Environment,” in *Proceedings of the 2020 9th International Connference on Software and Information Engineering*, 2020, pp. 174–179, doi: 10.1145/3436829.3436875.
[33] A. Troisi, “Ethological research in clinical psychiatry: the study of nonverbal behavior during interviews,” *Neurosci. Biobehav. Rev.*, vol. 23, no. 7, pp. 905–913, Nov. 1999, doi: 10.1016/S0149-7634(99)00024-X.
[34] E. G. Shreve, J. A. Harrigan, J. R. Kues, and D. K. Kagas, “Nonverbal Expressions of Anxiety in Physician-Patient Interactions,” *Psychiatry*, vol. 51, no. 4, pp. 378–384, 1988, doi: 10.1080/00332747.1988.11024414.
[35] C. Mohiyeddini, S. Bauer, and S. Semple, “Displacement Behaviour Is Associated with Reduced Stress Levels among Men but Not Women,” *PLoS One*, vol. 8, no. 2, pp. 1–9, 2013, doi: 10.1371/journal.pone.0056355.
[36] C. Mohiyeddini and S. Semple, “Displacement behaviour regulates the experience of stress in men,” *Int. J. Biol. Stress*, vol. 16, no. 2, pp. 163–171, 2013, doi: 10.3109/10253890.2012.707709.
[37] C. Mohiyeddini, S. Bauer, and S. Semple, “Stress The International Journal on the Biology of Stress Public self-consciousness moderates the link between displacement behaviour and experience of stress in women Public self-consciousness moderates the link between displacement behaviour and experie,” *Stress*, vol. 16, no. 4, pp. 384–392, 2013, doi: 10.3109/10253890.2012.755171.
[38] J. Team, “JASP: A Fresh Way to Do Machine Learning,” *JASP - Free and User-Friendly Statistical Software*, 2022. .
[39] D. Schmid and N. A. Stanton, “Exploring Bayesian analyses of a small-sample-size factorial design in human systems integration: the effects of pilot incapacitation,” *Human-Intelligent Syst. Integr.*, vol. 1, no. 2, pp. 71–88, 2019.
[40] Z. Dienes, “Using Bayes to get the most out of non-significant results,” *Front. Psychol.*, vol. 0, p. 781, Jul. 2014, doi: 10.3389/FPSYG.2014.00781.
[41] S. M. Merritt and D. R. Ilgen, “Not all trust is created equal: Dispositional and history-based trust in human-automation interactions,” *Hum. Factors*, vol. 50, no. 2, pp. 194–210, Apr. 2008, doi: 10.1518/001872008X288574.
[42] J. Dang, K. M. King, and M. Inzlicht, “Why Are Self-Report and Behavioral Measures Weakly Correlated?,” *Trends in Cognitive Sciences*, vol. 24, no. 4. Elsevier Ltd, pp. 267–269, Apr. 2020, doi: 10.1016/j.tics.2020.01.007.
[43] A. Troisi, “Displacement activities as a behavioral measure of stress in nonhuman primates and human subjects,” *Stress*, vol. 5, no. 1, pp. 47–54, 2002, doi: 10.1080/102538902900012378.