


RESEARCH ARTICLE OPEN ACCESS

Effect of Reliability, Its Framing and Error Bias on Trust in Human-Vehicle Collaboration

Jue Li¹  | Yilu Ye² | Long Liu¹ | Andreas Butz² 

¹College of Design and Innovation, Tongji University, Shanghai, China | ²Department for Informatics, LMU, Munich, Germany

Correspondence: Andreas Butz (butz@ifi.lmu.de)

Received: 30 May 2023 | **Revised:** 23 September 2024 | **Accepted:** 1 April 2025

Funding: Jue Li's contributions were funded by the China Scholarship Council (grant number: 202106260052).

Keywords: automated driving system | automation reliability | error bias | framing effect | human-vehicle collaboration

ABSTRACT

System reliability promotes trust, but may also impair human monitoring performance and in turn affects trust. This effect varies across different errors. This study examined the effect of automation reliability (100%, 75%, and 50%) and its framing (negative and positive description of reliability), and error bias (false alarm and miss) on user trust and its related factors in the automated driving system (ADS). Each participant completed 16 trials with human-vehicle collaboration task in a static driving simulator. The results showed that ADS with higher reliability positively impact user trust, but negatively impact situation awareness. Users' trust was higher in false alarm (FA) events than in miss events, but task success and situation awareness were higher in miss events. This study revealed an unusual negative correlation between trust and situational awareness in human-vehicle collaboration and provided possible insights into the internal factors of error bias in automation. Our finding has implications for reliability disclosure strategies and trust calibration.

1 | Introduction

Automated technologies involved in the vehicle industry is an inevitable trend for the future, as automated driving system (ADS) can reduce congestion, reduce driver's fatigue, improve road safety, and increase fuel efficiency (Daziano et al. 2017; Fagnant and Kockelman 2015). Sharing responsibility for vehicle control with the system also allows the driver to use the travel time for non-driving-related tasks (NDRTs) (Fagnant and Kockelman 2015). However, many users may harbor distrust based on their preconceived notions or news coverage of ADS failures (Shi et al. 2021), and it may subsequently lead to the disuse of part or all of the functions of the automation (Parasuraman and Riley 1997). Cases of ADS errors may affect user trust and use of ADS, and a key factor in this human-vehicle interaction is the automation reliability and the variables that may be associated with it (Mishler and Chen 2023).

The issue of reliability is of particular interest in the context of advanced automation. Previous studies have examined the effect of automation reliability on trust, suggesting that an increase in automation reliability leads to increased users' trust or reliance (Large et al. 2019; Mishler and Chen 2023), but may seriously impair their ability to monitor effectively, further affecting the human-automation trust relationship (Bailey and Scerbo 2007). And users are not sensitive to system reliability if there is no accurate description or feedback about it (Wang et al. 2009), especially when the system does not miss but makes false alarm, as false alarm (FA) and miss affect trust through two independent processes, and automation FAs may be more damaging than misses (Bliss and Acton 2003; Meyer 2001). An exploratory driving simulation experiment was conducted in this study, with several levels of reliability, description framings for reliability level, and error bias events simulating different driving scenarios to examine how user trust varies when collaborating with the ADS.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Human Factors and Ergonomics in Manufacturing & Service Industries* published by Wiley Periodicals LLC.

1.1 | Trust in Automation

Trust in automation refers to human trust in automation system, which depends on the performance, process, or purpose of an automated system (Lee and Moray 1992), and determines automation usage. Inappropriate trust in automation may lead to misuse, disuse, and abuse of automation (Parasuraman and Riley 1997). Lee and See (2004) arguably provide one of the most comprehensive descriptions to elucidate the role of trust in automation and propose a method for defining appropriate trust in automation which emphasizes that overtrust and distrust are all poor calibrations in which trust does not match the system capabilities. That means, with interaction, humans can understand the capabilities and limitations of the system and calibrate their trust accordingly (Parasuraman and Riley 1997).

The current study applied this concept of human trust in ADS, where the driver should trust that the vehicle will respond to unpredictable hazards exactly as it is programmed to do. Drivers can learn over time via interacting with ADS about its capabilities and calibrate their trust accordingly. In this process, automation errors are one of the most influential factors affecting trust, usually depending on error frequency, predictability, and severity. For minor automation errors, trust decrement might be small and inconsequential (Beggiato et al. 2015; Mishler and Chen 2023). Major or serious automation errors like those that cause vehicles to crash would be much more potent (de Visser et al. 2018). The decline of trust may prompt drivers to ignore or disable ADS if the driver trust does not recover from the decline, which is a direct consequence of mistrust and could hurt the performance of the entire human-vehicle team (Bliss and Acton 2003). Extensive evidence suggests that understanding the dimensions of trust in ADS will enable designers to create systems that maximize human-vehicle collaboration and thus improve road safety in the future.

How to measure trust in automation is also a key topic. In experimental studies in aviation automation, trust is usually measured through two behaviors: compliance (operator responds to an automation alert) and reliance (operator responds to an automation non-alert) (Chavaillaz et al. 2016; Dixon and Wickens 2006; Johnson et al. 2004). Such measurements can be employed in warning systems that allow human behavior to be evaluated in a single trial (Holthausen et al. 2020). In automated driving contexts, some of the driver behaviors in ADS cannot be simply classified as whether or not the driver is responding to the automation alerts. In some empirical studies that examined driver trust in ADS, physiological (such as pupil dilation (Stapel et al. 2022), galvanic skin response (Morris et al. 2017), and eye-tracking (Hergeth et al. 2016)) or psychological (such as variability of mental model (Beggiato et al. 2015)) measurements have been proposed, leading to diverse results. It remains unclear how valid physiology or psychology is over a longer period of time, aggregated over repeated events. There are also studies that used situation awareness (SA) and perceived reliability (PR) as proximate variables of trust (He et al. 2022; Petersen et al. 2019; Washburn et al. 2020). Ultimately, the most direct way to measure trust in ADS is by asking participants via a standardized or adapted scale. An often-used scale is the

“Trust in Automation Scale” proposed by Jian et al. (2000), which classified trust into competency, reliability, predictability, and faith. Cramer et al. (2008) proposed a scale containing 12 items, six of which were adapted from Jian et al. (2000). On the basis of these, Holthausen et al. (2020) developed the “Situational Trust Scale for Automated Driving”, which incorporates important elements of Hoff and Bashir (2015)’s model, such as the driving context’s potential risks and benefits, or the driver’s self-efficacy for operating an automated vehicle. In general, one possibility to evaluate trust in a specific human-vehicle interaction context might be the integrating measurements of different dimensions, by analyzing driving tasks and scenarios.

1.2 | Reliability, Error Bias and Framing

Through a meta-analysis of factors affecting trust in automation, Hancock et al. (2011) found that reliability is one of the main factors influencing the development of human trust. Azevedo-Sa et al. (2021) defined reliability as internal risk and demonstrated that reliability could influence user trust significantly because risk in driving has been discovered to determine whether trust translates into actual trusting behaviors. Most work in human-vehicle collaboration has shown that when the ADS is not very reliable, human collaborators need to react to and correct malfunctions that occur, and then their trust will decrease and they rely less on the system (Bailey and Scerbo 2007; Dixon et al. 2007; Dzindolet et al. 2003) across a range of driving automation context, including the entire ADS (Large et al. 2019), crash avoidance systems (Bliss and Acton 2003), and in-vehicle navigation (R. Ma and Kaber 2007).

When automation fails, it usually produces one of two types of errors. A FA is an incorrect indication of an event, while a miss means that an error is not detected (Dixon et al. 2006), collectively referred to as error bias. Previously, it had been assumed that FAs would hurt overall performance more than did misses, as well as affecting operator compliance and reliance (Dixon et al. 2007), and persistent or pervasive FAs would lead to lower operator trust in automation (Dixon et al. 2007; Johnson et al. 2004). FA are also prone to causing annoyance, which requires a higher workload and some otherwise unnecessary actions from human collaborators (Dixon and Wickens 2006). In fact, these phenomena are all intrinsically linked. Compliance and reliance as externally visible responses to trust, and the main difference between them in human-automation interactions is the absence of cues (Chancey et al. 2017). FAs are usually accompanied by perceptually salient events, that is, more noticeable or memorable errors than misses (Dixon et al. 2007), and also need more intervention from collaborators, thus, causing a lower trust in the automated system and true alarms are ignored, known as the “cry wolf” syndrome (Breznitz 1984; Parasuraman and Riley 1997). Despite these evidence, it is important to note that experts may be less receptive to misses than FAs, as the costs associated with missing a critical event in a signaling system can be disastrous (Chancey et al. 2017; Masalonis and Parasuraman 1999). In human-vehicle collaboration, the impact of these two error biases remains to be examined.

User trust is also shaped by any information they receive before human-automation interaction about the possible error or the cost of a failure (Groom et al. 2011; Wang et al. 2009). This information about the reliability of automation is referred to as framing (Hallahan 1999). Framing helps people negotiate and interpret interactions appropriately by encouraging the activation of knowledge structures, or relevant schema (Tannen and Wallat 1987), and thus accurately match their trust to the actual level of reliability in the subsequent interactions (Wang et al. 2009). It has been demonstrated that automation with framing generated higher evaluations than those without (Groom et al. 2011; Wang et al. 2009). High framing that describes the automation as high capabilities generated more positive evaluations than low framing (Paepcke and Takayama 2010; Washburn et al. 2020). Notably, Framing effects also influence human attitudes (Tversky and Kahneman 1981) - i.e. presenting the same message in a positive or negative way, people are usually more sensitive to a negative one, because the negative description suggests possible losses, while a positive description suggest possible benefit (Vliegenthart 2012; C. Zhang et al. 2022). This difference will be more pronounced in high-risk tasks (Tversky and Kahneman 1981), such as human-vehicle cooperation for autonomous driving. Because a negative framing implies the loss of a possible collision, while a positive framing implies the benefit of just a safe drive.

1.3 | The Current Study

Previous studies have generally suggested that the reliability of ADS has a significant impact on human trust and that error bias is a key factor in human-automation collaboration, but none of them focus on the framing effect in human-vehicle interaction and adopted these variables as independent variables to examine the effects of them and analyze their internal factors on user trust in human-vehicle collaboration. The current study addresses the gaps in knowledge about automation reliability and its framing, as well as error bias in driving systems, and puts a special focus on their interplay by examining trust development. The primary research question of the current study was: *How do automated driving system reliability and its framing, and error bias affect user trust?* Therefore, we designed and conducted a simulated driving experiment to investigate the effects of these three variables on participants' trust in a simulated driving scenario with human-vehicle collaboration task. To explore the effects clearly, we excluded variables unrelated to the collaboration task in the driving scenario to mitigate participants' cognitive workload. The experiment was designed with a simple scenario and collaboration task, involving driving on a low-volume road and encountering a moving obstacle with unknown intention upon arrival at an intersection.

In experiment studies of human-vehicle collaboration tasks, usually, if the system handles the driving task well, the driver is encouraged to engage in NDRTs (Radlmayr et al. 2014). When a complex situation arises, the system provides certain information to the driver, maintains the driver's situational awareness, or seeks the driver's intervention (Guo et al. 2018; Xing et al. 2021). In the current study, the basic form of human-vehicle

collaboration was similar to the predecessors. the effects were examined in specific driving scenarios, encompassing secondary tasks of NDRTs, and primary tasks of human-vehicle collaboration. Participants' performance in the collaboration task was considered to be an intuitive outward sign and objective measure of trust in the ADS. Also, the scales that participants were asked to fill out at the end of each collaboration task were used as subjective measurements of trust.

2 | Materials and Methods

2.1 | Participants

A-priori sample size calculation was performed using G*Power software (Version 3.1.9.7) (Faul et al. 2007). For an assumed power of 0.80, alpha of 0.05, and effect size of 0.25, we had a projected sample size of 30 participants. Therefore, the study recruited 30 participants from two universities, of which 15 self-identified as female and 15 as male. 28 participants were students (93%) and 2 employees (7%). The ages ranged from 20 to 31 years old, with a mean of 24.7 years (SD = 8.13). All participants reported having normal or corrected-to-normal hearing and vision. No participants had previously experienced driving simulators. They signed a consent and GDPR form before they took part in the study and were compensated with a 10-euro shopping voucher after the study.

2.2 | Apparatus and Stimuli

The experiments were conducted in Germany at the Department for Informatics of LMU, using a fixed-base open driving simulator with three 43-inches display monitors (3840 × 2160 resolution) providing a horizontal field-of-view of approximately 135°, which display the main driving scenario. The sampling frequency of the driving simulator is 60 Hz. The steering wheel, pedals, and gear stick during the experiment were covered, to eliminate the participants' sense of driving a manually driven car (See Figure 1). Participants sat approximately 120 cm from the primary monitors. A surround audio system provided the sound of the engine and surrounding traffic object in the scenario, which was placed under the three monitors.

The driving scenario of highly ADS in this study was simulated in Unity 3D (Version 2018.4.14f1) and run on all three monitors of the driving simulator. In the scenario, there was a 2-lane urban road with low traffic, about 1.2 km long with an intersection at the end of the road. Low visibility in foggy conditions has been found to increase the likelihood of collisions (Yan et al. 2014). To improve the plausibility of the vehicle system failures, we set the weather environment to a foggy day that permitted drivers to spot an obstacle 130 m away. Accordingly, the automated driving speed (about 60 km/h) was set slightly below the speed limit of many urban roads. So, the participants were able to identify the obstacle that could be reached after about 7 s (i.e., time limit for collaboration tasks on center console display) at this speed. It took about 65 s for the vehicle to drive 1.2 km and reach the intersection.

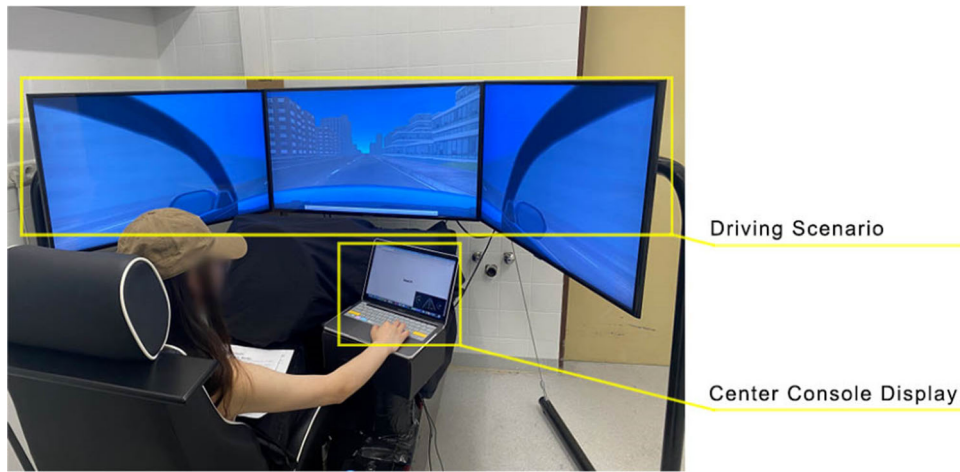


FIGURE 1 | Experiment environment and setup.

Within the experiment, the participants were seated in the simulator. On the right-hand side, an additional screen (Mac Book Pro, 13-inch, M1, 2020) was mounted to simulate the vehicle center console display, allow participants to participate in NDRTs, and act as interfaces for collaboration tasks, and was also used to fill in the scales at the end of each trial. The collaborative tasks and NDRTs were created in a web program created with JavaScript, where participants can interact with the touchpad and a keyboard.

2.3 | Experimental Tasks

For the participants, each trial in the experiment consisted of two tasks. The primary collaboration task was that when the vehicle was about to reach the intersection, the participant would see the vehicle's status interface zoomed in on the additional screen (see NDRTs to C in Figure 2). This interface showed a bird's eye view of the vehicle the participant was riding in, and provided the following information: road conditions, current charge/gasoline level, current speed, and collaboration task information. The collaboration task pop-up in the middle was translucent, with system detecting information, driving suggestions, and task countdown. The participant could choose whether to go straight or to brake by pressing the "up" or "down" key on the keyboard based on the suggestion on the pop-up and their own judgment of the scenario.

It is worth noting that unlike the signal system, which only alerts when it detects an error, the ADS in this study provides detecting information and driving suggestion on every trial. And the information and suggestion were always provided accordingly. That is, If the system detects correctly, it will give the correct information and the appropriate driving suggestion. Figure 3 depicts the driving scenarios and the vehicle's status interfaces when the system is reliable or unreliable for two error bias events. Regardless of whether the system information was correct, the collaboration task would be successful as long as the participant's response was correct. If the collaboration task succeeds, in FA events, the vehicle would brake at a certain distance from the pedestrian; in miss events, the vehicle would drive straight through the empty intersection. If failed, in FA events, the vehicle would drive into the intersection and have a

minor collision with the pedestrian; in miss events, the vehicle would brake and stop at the empty intersection for no reason.

The secondary task of NDRT in this study was a word game, modified from Jarosch et al. (2019)'s study. Participants had to detect all words beginning with "p" whenever it was presented on the additional screen by pressing the "space" key, among the random words beginning with "b," "q," "p" and "d."

After each driving task, participants were asked to fill out three scales. At the end of all trials, participants would answer several questions in a short interview related to the experiment to complement some unclarified parts of the experiment. All scales and interview outline are presented in Supporting Information S1: Additional File 1.

2.4 | Experiment Design

A mixed $3 \times 2 \times 2$ design was employed in this study. The level of reliability was a between-subject factor (high, medium, low). Error bias event (FA and miss) and framing (positive and negative) were within-subject factors that appeared in a counterbalanced order in each group. Each participant experienced 16 trials, with 4 repetitions of the crossover of framing conditions and error bias event conditions.

System reliability in the experiment was manipulated across the 3 experimental conditions by varying the percentage of correct detection and suggestions given by the ADS. Under high-, medium-, and low-reliability conditions, 100%, 75%, and 50% of the detection and suggestions were correct, respectively. Evidence suggests that the lower limit of acceptable automation reliability is 70%–80% (Johnson et al. 2004; Rovira et al. 2014). Thus, setting these three reliability conditions can represent perfect automation, acceptable automation, and unacceptable automation. Because reliability was a between-subject variable, 30 participants were randomly assigned to 1 of 3 reliability condition groups.

Error bias events refer to driving event that occurs when the participant is about to reach an intersection in the ADS. For FA events, there was no pedestrian at the intersection, but the system could potentially detect a moving obstacle ahead and

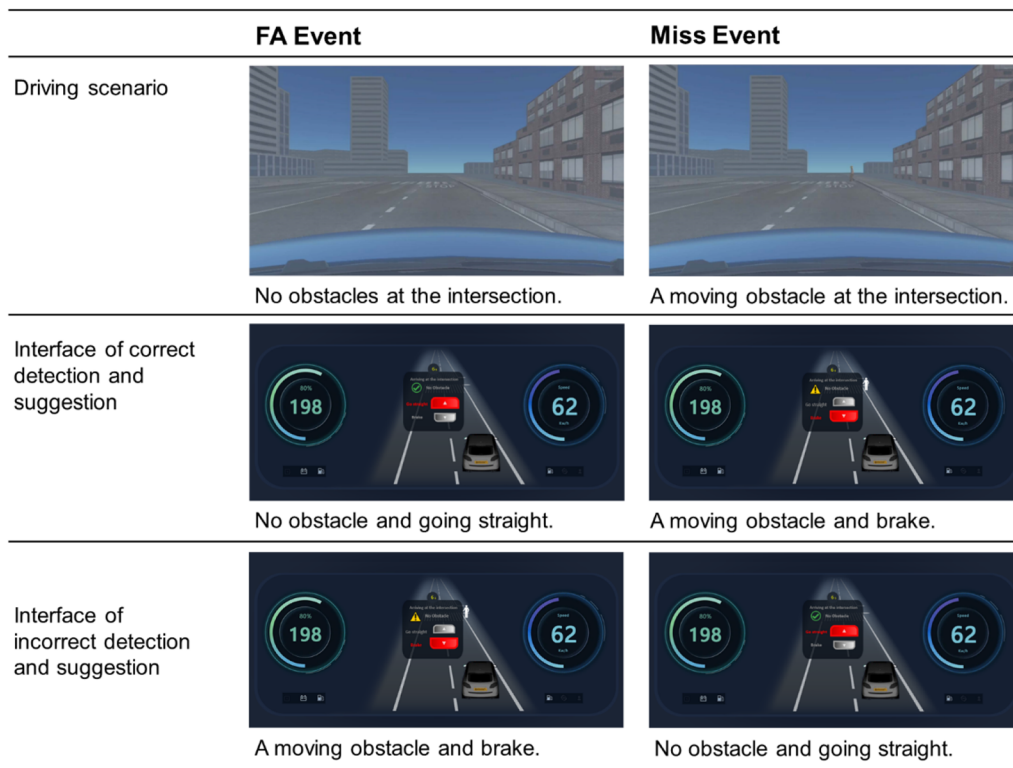


FIGURE 2 | Driving scenarios and the vehicle's status interfaces at the 2nd second of the collaboration task, up to 7 s.

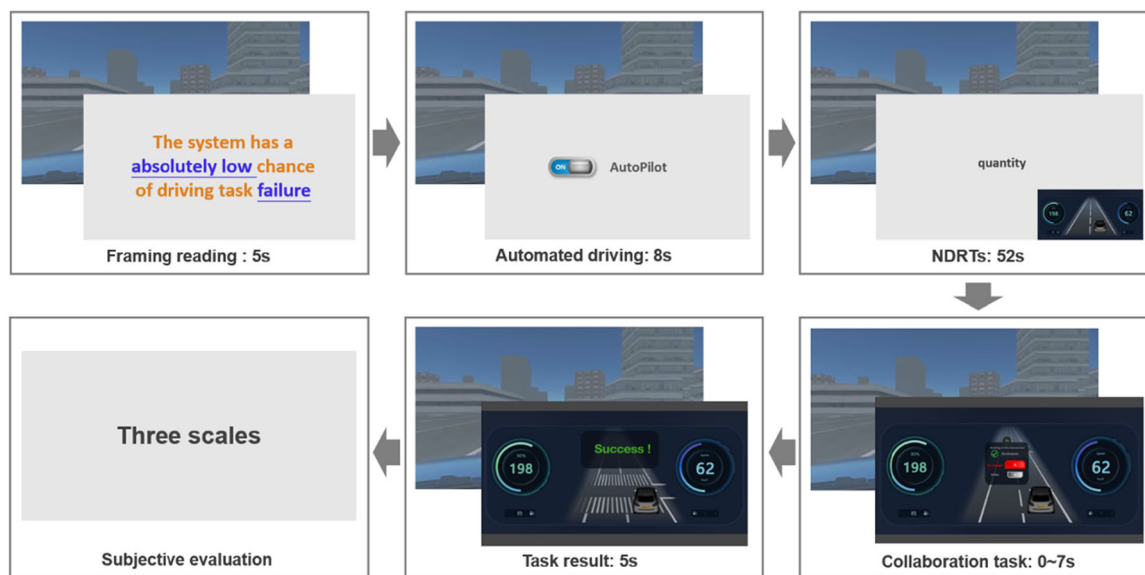


FIGURE 3 | Procedure for each trial. The front layer of each step are vehicle center console interfaces in the additional screen; the back layer are the driving scenarios in the driving simulator monitors.

suggest braking. For Miss events, there was a pedestrian on the side of the road about to cross the road, but the system could potentially detect no obstacle and suggest going straight. See Figure 3.

Framing in this study was a description of ADS reliability, presented on the additional screen at the beginning of each trial, with two conditions of positive and negative, describing the chances of success or failure in the primary task. Evidence suggests that consumer trust in new technologies stems from a

balance of perceived risks and perceived benefits (Ali et al. 2021; Featherman et al. 2021). The positive framing implies the chance that the participant will drive safely through the intersection, while the negative framing implies the chance that the participant will collide with a pedestrian at the intersection. See Table 1 for further information about framing under three reliability conditions.

Both error bias and framing were within-subject variables, so participants were required to participate in all four conditions

TABLE 1 | Framing under three reliability conditions.

System reliability	Positive framing	Negative framing
High	The system has an absolutely high chance of driving task success.	The system has an absolutely low chance of driving task failure.
Medium	The system has a relatively high chance of driving task success.	The system has a relatively low chance of driving task failure.
Low	The system has a low chance of driving task success.	The system has a high chance of driving task failure.

for the two variables. To eliminate the effect of trial sequence, 3 groups of 16 simulated driving trials were balanced by a Latin square, and then 10 sets of sequences were randomly selected from 16 sets of sequences. Finally, 3 groups of 10 sets of different trial sequences were obtained.

2.5 | Dependent Variables

During all drives, the behavior measurements of participants' performance during the collaboration task and the subjective measurements filled in a series of scales have been obtained as dependent variables. Participants' performance including compliance, response time (RT), and task success in the collaboration task. Compliance was defined as whether participants responded the same as the system suggested in the collaboration tasks; RT refers to the time for participants' first response to the collaboration task. Collaboration task success refers to whether the vehicle brakes at a certain distance from the pedestrian or drives straight through the empty intersection.

The subjective measurements included trust, perceived reliability (PR), and situation awareness (SA). Trust was measured with the *Situational Trust Scale for Automated Driving (STS-AD)* which comprises 6 items: trust, performance, NDRT, risk, judgment, and reaction (Holthausen et al. 2020). PR scale was adapted scale from Washburn et al. (2020) *Reliability Questionnaire* and Kidd and David (2003) *Subscale of Reliability*, including 5 questions (see Table 1). *Adapted Situational Awareness Rating Technique (SART)* Taylor (1990) was used to measure participants' SA, consisting of 9 questions in dimensions of instability, complexity, variability of situation, arousal, concentration, division of attention, Spare Mental Capacity, information quantity, and familiarity with the situation. Those three scales were using a 5-point Likert scale which ranged from 1 to 5 with "Strongly Disagree" to "Strongly Agree" or "Very Low" to "Very High".

2.6 | Procedure

Upon arrival at the laboratory, each participant was told the primary aim of the study and was asked to complete a consent form and a demographic sheet. They were asked to sit in the simulator and encouraged to adjust the vehicle's chair to comfortably reach the additional screen. Then the participants had a task introductory and practice session that lasted approximately 8 min with 2–3 trials and covered the two error bias event conditions to practice and get familiar with the tasks. During this session, participants were told that they should treat the

experiment as if they were driving on real roads and complete primary and secondary tasks as best they could.

After the practice sessions, a series of 16 formal trials began. For each trial, participants were tasked with both driving collaboration task and performing the NDRTs. First, a framing of system reliability was presented on the additional screen through positive or negative text. This text framing would last for 5 s and the adverbs and adjectives for reliability level in the text were highlighted to ensure that participants captured the information. Then the autopilot was turned on. Participants could enjoy a few seconds driving through the monitors of the simulator and see the vehicle driving automatically on a foggy urban road. After a beep, participants should join the NDRTs on the additional screen for 52 s until the vehicle almost reached an intersection and the vehicle's status interface popped up on the additional screen replacing the NDRTs interface. Participants were asked to choose to comply or not comply with the suggestion within 7 s. After driving, participants completed three scales based on the drive just finished. Figure 2 illustrates the monitors of the driving scenario and interfaces on the additional screen that participants can see throughout the trial procedure. Between each two trials, it was reiterated that the participants will now experience a different ADS from a different supplier. Such a design was implemented to eliminate the effects of framing and results from the previous trial. Also, participants were asked to take a 1- or 2-min break before moving on to the next trials.

At the end of all 16 trials, participants would answer questions in a short interview. And then the participation fee was delivered and the experiment ended. The total duration of the experimental procedure consisted of four sessions and was approximately 80 min (see Figure 4).

3 | Results

The experimental data of 30 participants were summarized, including participants' performance and scale data after driving. As the scale data collected in the experiment was from subjective measurements, Cronbach's alpha firstly was used to determine the degree of consistency and reliability for Trust, PR, and SA scales. The calculated values of Cronbach's alpha for Trust, PR, and SA were 0.809, 0.865, and 0.785, respectively, indicating that all responses were reliable.

Given that the study aimed to analyze the relationships between variables, statistical analyses were conducted using correlational approaches. The Shapiro-Wilk Test and Q-Q Plot were first used

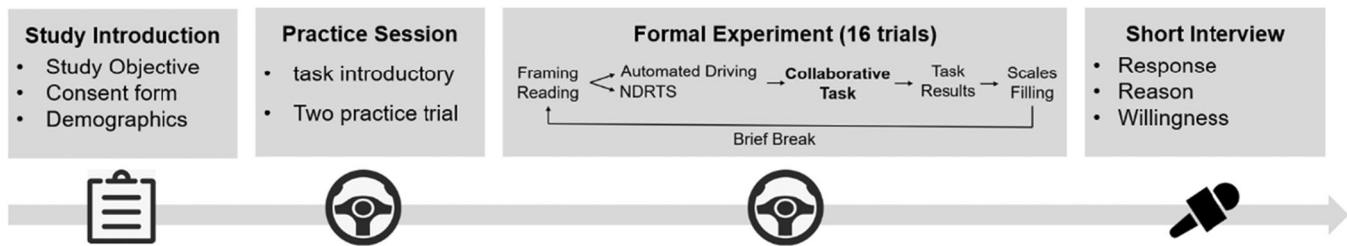


FIGURE 4 | Experiment procedure.

to check the normality of all dependent variables, as the experimental sample size did not exceed 100 (Zhang and Wu 2005). Dependent variables that were normally distributed and met the homogeneity of variance assumption (Levene's test) were analyzed using ANOVA. For non-normally distributed data, discrete and continuous variables were analyzed using the Mann-Whitney test (two condition variable) and Kruskal-Wallis test (three condition variable). Binary Logistic Regression was used to predict binary variables. The alpha level was set to 0.05 for all statistical tests. All the analyses were processed using SPSS Statistics.

Regarding the post-experimental interview, a thematic analysis approach adapted from Braun and Clarke (2006) was chosen as the methodology in our post-experimental interview. The first author performed the coding process, based on the transcripts' semantic content, using raw quotes as codes. The codes were sorted into thematic categories based on repetition, similarities, and differences (Ryan and Bernard 2003). Within each thematic category, the codes were further differentiated and sorted into sub-themes. Then the second author reviewed the definitions and the names of the themes and gave feedback regarding the analysis, and the authors conjointly decided on the revisions.

3.1 | Descriptive Statistics

The descriptive statistics for the combination groups of the levels of each dependent variable are shown in Table 2. From the descriptive statistics, the three measurements of participants' performance, Trust score, and PR score, showed high consistency. Because of the simplicity of the collaboration task, the compliance rate and task success rate were 100% in several groups, mainly in high-reliability conditions. The shortest RT ($M = 0.875$, $SD = 0.607$) for the collaboration task was in the high-reliability* miss* negative group. The highest Trust ($M = 4.19$, $SD = 0.393$) and PR score ($M = 4.38$, $SD = 0.510$) were in the high-reliability* FA* positive group. The lowest compliance rate ($M = 50$, $SD = 0.506$) and lowest Trust score ($M = 2.97$, $SD = 0.614$) were in the low-reliability* miss* positive group. The longest RT ($M = 1.88$, $SD = 0.939$) was in the low-reliability* FA* negative group. The lowest success rate ($M = 88$, $SD = 0.420$) was in the low-reliability* FA* positive group. The lowest PR score ($M = 2.89$, $SD = 0.770$) was in the low-reliability* miss* negative group.

SA scores showed different results from other dependent variables. The highest SA scores ($M = 3.54$, $SD = 0.565$) were in the

low-reliability* miss* negative group and the lowest ($M = 2.55$, $SD = 0.636$) were in the high-reliability* FA* negative group.

3.2 | Collaboration Task Performance

Collaboration task performance included participants' compliance, response time (RT), and task success, which were behavioral indicators of participant trust in the ADS. Higher compliance rates and shorter RT mean that participants were more willing to follow the system's suggestion in the collaboration tasks, while higher task success rates indicate higher collaboration quality.

Participants' compliance data was analyzed using a Binary Logistic Regression to ascertain the effects of system reliability, error bias, and framing on the likelihood that a participant would follow the suggestion of the ADS. The logistic regression model was statistically significant, $\chi^2(4) = 131.84$, $p < 0.001$. The model explained 36.0% (Nagelkerke R^2) of the variance in compliance and correctly classified 76.0% of cases. Hosmer and Lemeshow test showed a good fitting of the logistic regression model, $\chi^2(8) = 0.237$, $p > 0.05$. Participants in the medium-reliability condition were 2.98 times more likely to follow the suggestion of the ADS than in the low-reliability condition, Wald $\chi^2(1) = 20.017$, $p < 0.05$. There were lower compliance rates for the low-reliability condition ($M = 52.0\%$, $SD = 0.501$) compared to the medium-reliability condition ($M = 76.0\%$, $SD = 0.427$) and high-reliability condition ($M = 100.0\%$, $SD = 0.000$). The effect of error bias and framing on participants' compliance was not statistically significant, $p > 0.05$.

RT data revealed no outliers but were not normally distributed, $D(480) = 0.793$, $p < 0.05$. Tests showed a significant effect of reliability on RT, $H(2) = 49.53$, $p < 0.05$, and a significant difference between the high-reliability condition (mean rank = 192.24), medium-reliability condition (mean rank = 241.15), and low-reliability condition (mean rank = 288.11), $p < 0.05$ for all comparisons, suggesting that the RT for human-vehicle collaboration task was shorter with higher reliability. The effects of error bias and framing on participants' RT were not statistically significant, $p > 0.05$.

Task success data was analyzed using a Binary Logistic Regression to ascertain the effects of the independent variables on the likelihood that a participant would fail in the collaboration tasks. The logistic regression model was statistically significant, $\chi^2(4) = 23.304$, $p < 0.001$. The model explained 22.7% (Nagelkerke R^2) of the variance of task success, and correctly classified 97.5% of cases. Hosmer and Lemeshow test

TABLE 2 | Descriptive statistics of the dependent variables.

c	Compliance rate (%)		RT		Success rate		Trust		PR		SA	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
HR* FA* NF	100	0.00	.975	0.48	100	0.00	4.00	0.41	4.30	0.54	2.55	0.64
HR* miss* NF	100	0.00	.875	0.61	100	0.00	4.03	0.46	4.28	0.53	2.91	0.63
HR* FA* PF	100	0.00	1.20	0.56	100	0.00	4.19	0.39	4.38	0.51	2.56	0.67
HR* miss* PF	100	0.00	1.00	0.50	100	0.00	3.98	0.41	4.26	0.60	2.92	0.58
MR* FA* NF	80.0	0.41	1.42	0.63	95.0	0.22	3.55	0.41	3.82	0.65	2.73	0.39
MR* miss* NF	75.0	0.44	1.05	0.75	100	0.00	3.42	0.38	3.60	0.79	3.30	0.40
MR* FA* PF	75.0	0.44	1.48	0.75	98.0	0.16	3.40	0.57	3.73	0.77	2.84	0.37
MR* miss* PF	75.0	0.44	1.15	0.62	100	0.00	3.22	0.68	3.58	0.76	3.33	0.33
LR* FA* NF	53.0	0.51	1.88	0.94	93.0	0.27	3.17	0.556	3.07	0.66	3.02	0.49
LR* miss* NF	53.0	0.51	1.32	0.73	98.0	0.16	3.15	0.63	2.89	0.77	3.54	0.57
LR* FA* PF	53.0	0.51	1.75	0.67	88.0	0.34	3.08	0.64	3.22	0.762	3.02	0.48
LR* miss* PF	50.0	0.51	1.35	0.53	100	0.00	2.97	0.61	2.92	0.843	3.47	0.65
Total	76.0	0.43	1.29	0.72	98.0	0.16	3.51	0.66	3.67	0.866	3.02	0.61

Abbreviations: FA, FA events; HR, high-reliability; LR, low-reliability; Miss, miss events; MR, medium-reliability; NF, negative framing; PF, positive framing.

showed a good fitting of the logistic regression model, Wald $\chi^2(5) = 0.885$, $p > 0.05$. Error bias was the only statistically significant variable, $\chi^2(1) = 5.55$, $p < 0.05$. The task success rate for participants in the miss event condition ($M = 100.0\%$, $SD = 0.065$) was 11.95 times higher than in the FA event condition ($M = 95.0\%$, $SD = 0.210$). Although the predictability of reliability was not statistically significant, $p > 0.05$, the task success rate for participants in the medium-reliability condition ($M = 98.0\%$, $SD = 0.136$) was 3.21 times than in the low-reliability condition ($M = 95.0\%$, $SD = 0.231$), $\chi^2(1) = 2.91$, $p = 0.088$. The effect of framing on task success was also not statistically significant, $p > 0.05$.

3.3 | Trust

A 6-item, 5-point Likert scale was used to measure participants' trust in the ADS after the collaboration tasks. Tests showed the data were approximately normally distributed, $D(160) = 0.987$, $p > 0.05$ in the low condition of reliability. Thus, parametric analyses were performed. A Three-Way ANOVA revealed a significant effect on trust of reliability, $F(2, 468) = 104.23$, $p < 0.05$, $\eta^2_p = 0.375$. Post Hoc tests showed significant differences between these three conditions of reliability, $p = 0.000$ for all comparisons. The average trust score was 4.05 ($SD = 0.609$), 3.40 ($SD = 0.531$), and 3.09 ($SD = 0.423$) for the high-, medium-, and low-reliability conditions, respectively. There was also a significant effect of error bias on trust, $F(1, 468) = 2.844$, $p < 0.05$, $\eta^2_p = 0.375$, indicating a higher trust in the FA event condition than in the miss event condition. The average Trust score for FA event and miss event condition was 3.57 ($SD = 0.647$) and 3.46 ($SD = 0.672$). Figure 5 depicts participants' trust in all conditions of reliability and error bias. There was no significant effect of framing on Trust, $F(2, 468) = 2.844$, $p = 0.092$, $\eta^2_p = 0.006$, although the average trust score in negative framing condition ($M = 3.56$, $SD = 0.596$) was numerically higher than in positive framing condition ($M = 3.47$,

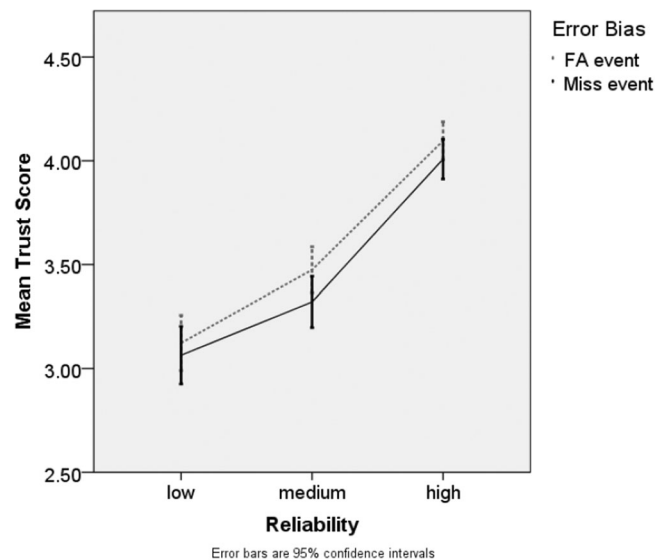


FIGURE 5 | Trust score for all conditions of reliability. Each line represents a different error bias events condition. Error bars are 95% confidence intervals.

$SD = 0.718$). The difference for the interaction comparisons was not significant, $p > 0.05$.

3.4 | Perceived Reliability

The PR scores were obtained by averaging the scores in the 5-item, 5-point Likert scale, which was not normally distributed, $D(480) = 0.952$, $p < 0.05$. Non-parametric analyses were applied. There was a significant effect of reliability on PR, $H(2) = 181.44$, $p < 0.05$, and a significant difference between high-reliability condition (mean rank = 344.70), medium-reliability condition (mean rank = 240.03), and low-reliability condition (mean rank = 136.78), $p < 0.05$ for all comparisons, indicating that

the PR score was higher in the ADS with higher reliability. Mann-Whitney tests revealed that there was no significant effect on participants' PR of error bias and framing, $p > 0.05$.

3.5 | Situation Awareness

Participants' SA scores were obtained by averaging the scores in the 9-item, 5-point SA scale. Tests revealed the data were approximately normally distributed, $D(160) = 0.985$, $p > 0.05$ in the medium-reliability condition. A Three-Way ANOVA showed that the main effect of reliability was significant, $F(2, 468) = 40.24$, $p < 0.05$, $\eta^2_p = 0.147$. Post Hoc Tests showed significant differences between the three conditions of reliability ($p < 0.05$ for all comparisons). The average SA score was 2.74 (SD = 0.648), 3.05 (SD = 0.457), and 3.26 (SD = 0.599) for the high-, medium-, and low-reliability conditions, respectively. There was also a significant effect of error bias on SA, $F(1, 468) = 90.111$, $p < 0.05$, $\eta^2_p = 0.161$, indicating a higher SA score in the miss event condition than in the FA event condition. The average SA score for the miss event and FA event conditions were 3.25 (SD = 0.588) and 2.79 (SD = 0.548). No significant effect of framing, and interaction effects on participants' SA were found, $p > 0.05$. SA scores of participants between groups of the statistically significant factors are reported in Figure 6.

3.6 | Post-Experimental Interview

We derived a total of 976 initial codes using descriptive inductive coding of the transcripts. After refining the removal of the duplicates, there were 660 codes left. While categorizing these codes based on their shared sense, we received 19 sub-themes containing more than 15 clustering codes. These sub-themes were then organized under 4 organizing themes: (1) Error, (2) Trust, (3) Human-machine interface (HMI) and (4) Safety.

Table 3 shows an overview of the seven themes and their respective sub-themes. The detailed interpretation of the results

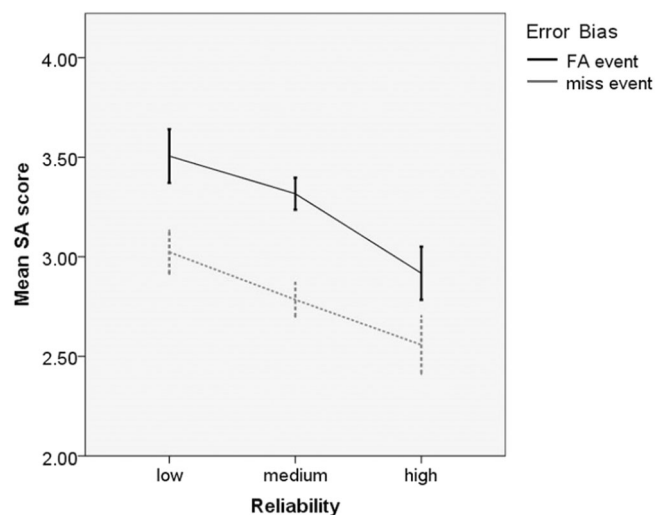


FIGURE 6 | SA for all condition of reliability. Each line represents a different error bias events condition. Error bars are 95% confidence intervals.

will be discussed along with the experimental results in the discussion section.

4 | Discussion

While it is important to continue improving ADS reliability, it is also critical that we learn how to support efficient human-vehicle teaming in the context of various degrees and types of error. The results in this study highlights that participant in the automation with higher reliability had a higher follow rate (i.e., compliance) as well as subjective trust evaluation (i.e., Trust), and PR on ADS, and also performed better, such as shorter RT in human-vehicle collaboration tasks, crossing different error bias events and framings, compared to lower reliability groups. It was expected because Hoff and Bashir (2015) model revealed that reliability is the primary internal factor influencing learned trust. However, the effect of reliability is not reflected in the indicator of task success. It's contrary to that of Zhou et al. (2022), who found that higher reliability improved performance in an identification human-automation collaboration task. One possible reason for it is that in our experiment, the collaboration task was to choose whether to brake or go straight within 7 s, which is much simpler than in a real-life traffic situation. In fact, participants successfully completed 98% of the tasks in all trials, which also suggested that participants took the collaborative task seriously and had high reliability of response. Although the task success data were not statistically significant, they still show a distribution with the highest success rate in the high-reliability condition (100%), and the lowest in the low-reliability condition (94%).

Unexpectedly, However, the distribution of the SA score was quite different from the other indicators. The finding also did not match previous findings by Petersen et al. (2019) and Miller et al. (2014) that SA and trust are positively related, due to the fact that higher SA allows drivers to better understand the environment and predict future actions, reducing uncertainty and thus enhancing trust. In our study, the highest SA score was in the low-reliability condition. From the study design of our experiment, the ADS always gave the correct detection and suggestion in high-reliability conditions, which may make the participant more dependent on the system and less observant of the driving environment, and in turn, impaired the participants' SA. According to one participant in the high-reliability group, she virtually stopped observing the environment and followed the system suggestions directly in the last few trials because of the vehicle's reliable performance. It confirmed what Dixon et al. (2007) suggested that highly reliable systems may severely impair an operator's ability to monitor for unanticipated system states, and a reduction in abilities of monitoring and prediction can further damage operator's SA. It is inferred that the contribution of understanding and prediction of the driving environment to participants' trust in ADS is actually insignificant, given the impact of apparent changes in system reliability on driver monitoring. Although many studies in the past have focused on SA in automated driving and have attempted to design vehicle HMIs to support driver SA (Kim et al. 2024; S. Ma et al. 2021), to enhance driver trust in the vehicle and driving safety (Parasuraman et al. 2008; Petersen et al. 2019). However, it can be asserted that the contribution of understanding and

TABLE 3 | Overarching themes and sub-themes.

Theme	Subtheme	Code examples
Error (52)	Alert (39)	False alarm; Unintelligent alert; incorrect reporting; incorrect cue
	Miss (30)	Missed detection; ignore obstacle; Missed pedestrian
	Vehicle control (30)	Braking; brake for no reason; go straight; decelerate; take over
	Detection (24)	Detection failure; misidentification; radar detection
	Find error (21)	Easier to find errors; notice the pedestrian; empty intersection
Trust (40)	Distrust (27)	Don't trust; never trust; untrustworthy vehicle; trust is irresponsible; not willing (to)
	Trust related behavior (27)	Not accept; rely on yourself; reliance on system judgment; re-judgment myself; obey the system; willing buy;
	Trust (25)	Tended to trust; fairly trust; trust in; willing
	Personality trait (19)	Uneasy; adventurous; enthusiast; pessimistic
Human-machine interface (37)	Framing perception (35)	No difference; more alarming; not impressed; unaware about
	Driving information (32)	obtain additional information; detailed information; concrete status of automated driving; send a message
	Functionality (24)	Display more; reminder light; explain the situation; bird's-eye view; how to interact
	Driving suggestion (23)	Correct suggestion; understand the suggestion; don't need suggestion;
Safety (37)	Ways of alert (16)	Sound of alert; Presented in a head-up display; more explanation pop-up; not visible enough; haptics; be coherent
	Risk (31)	Risk on road; risk of; involved in; take risk; dangerous; risky
	reliable (30)	Often makes mistakes; reliable vehicle; low-reliability; the system is uncertain
	Consequence (20)	Terrible consequences; frightened pedestrian; rear-end collision; accident; too severe
	Scenario (18)	on the real road; obstacle in the scenario; decision in particular situation; urgent situation; Traffic signal; red light
	Road users (15)	Pedestrian; cyclist; eye-contact; motion cues; other car;

prediction of the driving environment to participants' trust in ADS is actually insignificant, considering the of apparent changes in system reliability.

The effect of error bias on participants' SA was also statistically significant, with higher scores in the miss event condition than in the FA event condition. The research in SA is plentiful in the human-automation interaction domain, yet there is essentially none on how it pertains to error bias. The post-experimental interviews provided some possible arguments. One participant responded that she found both FA and miss error bias in those trials, and it is easier to find errors in miss events than in FA events. Considering further that participants' RT was shorter in miss events (1.12 s) than in FA events (1.45 s) (not statistically significant), it can be inferred that, it required more time and cognitive workload to find a FA error than a miss error in such driving events, which encouraged participants to better notice and process the information in FA events. Previous studies on error bias have made similar claims, that a FA-prone system may leave the operator less inclined to pay any attention to the entire automated domain (Dixon and Wickens 2006). The results on error bias and SA do not match the predictions and provide novel findings about the influences of reliability and error bias on participants' SA in ADS, revealing that regardless

of the relationship between SA and external variables, the determinants of SA remain its internal variables, that is, the level of attention and perception of the scenario.

The insights from SA data and interviews can also be used to explain the higher task success rate for the miss event condition compared to the FA event condition. Since it is easier for participants to detect errors by themselves in the miss events, the probability of correcting errors in miss events and succeeding in collaboration tasks is higher when both errors occur with the same frequency. It matches previous findings that FA-prone automation hurt performance more than miss-prone automation (Bliss and Acton 2003; Maltz and Shinar 2003). But the interpretations in these studies were primarily from the perspective of "cry wolf" syndrome (Breznitz 1984), in which FA alerts are more salient so that operators may ignore future warnings from the automation. It may not be a good explanation for lower success rate in the FA event condition of this study since the alert salience of the error bias has been manipulated to be the same level (visual detection information and driving suggestions, auditory and visual countdown). In fact, Rice and McCarley (2011) provided rare evidence that FAs produced poorer human performance and engendered lower automation use than misses even misses and FAs were matched

for their perceptual salience. They assert that automation misses and FAs may differ in their inherent cognitive salience, the degree to which they are noticed or remembered, independent of perceptual salience. The current data provide convincing evidence that FAs not only produce effects on participants' SA and RT in ADS that are qualitatively different from those produced by misses but also are quantitatively more harmful to task success rate than misses.

The results of Trust revealed a significant difference between error bias, with participants' rating Trust higher in the FA event condition than in the miss event condition. This is at odds with the assumption and previous studies that FAs will produce lower levels of subjective trust than misses in automated systems (Bahner et al. 2008; Bliss and Acton 2003; Dixon et al. 2007). Similarly, these studies typically attribute the sufficient distrust of FA to the "crying wolf" syndrome, yet the participants' levels of notice for FAs and misses in our experiment could be regarded as identical. There was also study suggested that misses would be more salient and may severely degrade trust when the system misses events that are easily detectable by the operator (Madhavan et al. 2006), as the costs associated with a signaling system's missing a critical event are potentially disastrous (Chancey et al. 2017). If participant and automation failed in the miss events in this study, the participant would face severe consequences, that is, a minor collision with a pedestrian. This cost of failure is much larger than an unwarranted stop at an empty intersection in the FA event. The interviews confirmed this speculation. Several participants mentioned that the consequences of a collision with a pedestrian in seemed too severe. And a participant believed that if the vehicle braked for no reason in FA events in a real driving scenario, it was possible to have a rear-end collision with the car behind him, and he might reconsider his trust strategy. Regardless, theories of trust, together with the data of trust and interviews in this study, confirm that the risk in system error likely plays a key role in determining trust in automation. Although it is unclear from our data how the consequences of failure under the two error biases, that is, the risks under the two conditions, influenced the participants' trust.

The data from this study also revealed inconsistencies in several performance indicators, that is, participants were quicker to collaborate with the ADS and more successful in the miss event condition, but preferred to follow the ADS suggestions in the FA event condition, although the effect of error bias on compliance rate was not statistically significant (77% in FA event condition, 75% in miss event condition). It suggested that one of the best behavioral indicators that represent the abstract concept of trust is compliance, that is, whether to follow the system. While other performance indicators are influenced by many factors. It can correspond to Mayer's definition of trust - the willingness to accept vulnerability (Mayer et al. 1995), due to which participants choose to trust the ADS's information and follow its suggestion in collaboration tasks.

From a comprehensive perspective, the performance and subjective measurements data in this study showed that error bias affects participants' trust and perception in human-vehicle interaction in two aspects: First, it's easier to find the system missing than the system making a FA, so participants

spent less time succeeding in the task, which resulted in lower SA. Second, as errors occur and tasks are failed, participants perceived greater risk in miss events than in FA events and in turn are less likely to follow and trust the ADS in those following trials. Whether this subjective and behavioral distrust will further negatively impact SA and task performance needs to be examined in a longitudinal human-vehicle collaboration.

Finally, when discussing the positive and negative aspects of framing, the data from this study indicated that framing did not affect any of the variables in terms of data, inferring framing effects were not reflected in human-vehicle collaboration. Several participants across both framing conditions reported that they did not find the differences between the two framing, and one of them noted that there were too many trials in the experiment that by the last few drives he was somewhat numb to the framing. Another participant said that he wanted to understand the specific degree of adverbs in the framing, and even tried to predict the pattern of the framing's emergence. The gap in research on the framing effect in automation leaves this result lacking in interpretable space. More research needs to be done to determine whether the framing effect, which has had a huge impact in the sociological and psychological fields, is unrelated to user trust and performance in human-automation collaboration.

4.1 | Limitations and Future Research

The current study used multiple measurements to examine trust in human-vehicle collaboration. Although the measurements are all related to trust, the development of some indicators varies considerably from trust in specific situations, resulting in inconsistencies in the variable. Future research on trust in automation should consider more behavioral indicators as dependent variables, such as dependence, compliance, and obedience. Regarding subjective data, larger sample size and more precise Likert-points might improve the data's normal distributivity, thus meeting the prerequisites for parametric tests to find more associations from the data, for example, the interaction effect between variables.

Another limitation is that error bias presented in the ADS in our study was two events, and was coarse-grained classification. When error bias was introduced to automated driving across domains, it had to be placed in specific driving events, resulting in differences in the error bias events in terms of salience of alert, cost of failure, and difficulty of humans to detect the error. As discussed in the previous chapter, it may not be the error bias itself that affects user trust, but rather the finer elements of it. Coarse-grained examination of error bias in human-automation may lead to many contradictory conclusions, such as that user trust is sometimes high in FA-prone systems and sometimes high in miss-prone systems. Thus, future work on error bias in human-vehicle interaction could attempt to delineate the intrinsic dimensions of error bias or to consider designing framing to describe the errors and aim to establish mechanism that how error bias acts on human performance. Systems that are prone to FA and miss also should be studied differently, as in our study not only the behavioral indicators

are different, but also the inherent psychological schemata that drive these behaviors.

Finally, the human-vehicle collaboration task in our experiment was set in a static driving simulator and only one vehicle driven by the participant was on the simulated urban road. While this manipulation could reduce risk and facilitate control of variables, leading to a lack of risk perception, and further reducing human sensitivity to perceived reliability and situational awareness that participants should have in the real world (Walker et al. 2018; Xu et al. 2018). This may be one of the reasons why some of the effects on subjective data in this study were not significant. In the future, we would like to refine these setups in a simulated experiment or test in multi-lane dynamic scenarios with test vehicles.

5 | Conclusion

In this study, we experimentally manipulated the level of system reliability, error bias of driving events, and framing of the reliability for investigating participants' user performance, situational trust, perceived reliability, and situation awareness in human-vehicle collaboration. Our findings contribute to the research on the antecedents of the effect of reliability on trust in automation, suggesting that the level of automation reliability has an important impact on drivers' performance, perceived reliability, and trust in the vehicle. However, users' situation awareness in a system with higher reliability was lower, attributed to the fact that the high-reliability and the positive framing of the system increases the participants' reliance and thus reduces their monitoring performance of the scenario. These findings have practical implications for reliability disclosure strategies of automated driving systems.

We also found that user trust is higher in systems prone to FA than miss. One possible explanation is that this study eliminated the difference in the salience of alert between FA and miss, and the other is that participants were shown the consequence of task failure after collaboration task, allowing them to perceive a higher potential risk in miss events than in FA events. The findings on error bias provide an in-depth discussion of the internal variables differences between automation FAs and misses, thereby allowing generalization of its implications beyond systems where only FA has salient alerts.

Acknowledgments

The authors would like to thank the participants in this study for their time and effort, and reviewers for their constructive and insightful feedback. Jue Li's contributions were funded by the China Scholarship Council (grant number: 202106260052). Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Ali, S., M. A. Nawaz, M. Ghuffran, S. N. Hussain, and A. S. Hussein Mohammed. 2021. "GM Trust Shaped by Trust Determinants With the Impact of Risk/Benefit Framework: The Contingent Role of Food Technology Neophobia." *GM Crops & Food* 12, no. 1: 170–191. <https://doi.org/10.1080/21645698.2020.1848230>.
- Azevedo-Sa, H., H. Zhao, C. Esterwood, X. J. Yang, D. M. Tilbury, and L. P. Robert. 2021. "How Internal and External Risks Affect the Relationships Between Trust and Driver Behavior in Automated Driving Systems." *Transportation Research Part C: Emerging Technologies* 123: 102973. <https://doi.org/10.1016/j.trc.2021.102973>.
- Bahner, J. E., A.-D. Hüper, and D. Manzey. 2008. "Misuse of Automated Decision Aids: Complacency, Automation Bias and the Impact of Training Experience." *International Journal of Human-Computer Studies* 66, no. 9: 688–699.
- Bailey, N. R., and M. W. Scerbo. 2007. "Automation-Induced Complacency for Monitoring Highly Reliable Systems: The Role of Task Complexity, System Experience, and Operator Trust." *Theoretical Issues in Ergonomics Science* 8: 321–348.
- Beggiato, M., M. Pereira, T. Petzoldt, and J. Krems. 2015. "Learning and Development of Trust, Acceptance and the Mental Model of ACC. A Longitudinal On-Road Study." *Transportation Research Part F: Traffic Psychology and Behaviour* 35: 75–84.
- Bliss, J. P., and S. A. Acton. 2003. "Alarm Mistrust in Automobiles: How Collision Alarm Reliability Affects Driving." *Applied Ergonomics* 34, no. 6: 499–509. <https://doi.org/10.1016/j.apergo.2003.07.003>.
- Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3, no. 2: 77–101. <https://doi.org/10.1191/1478088706qp063oa>.
- Brenzitz, S. 1984. Cry Wolf: The Psychology of False Alarms. https://xueshu.baidu.com/usercenter/paper/show?paperid=5d04a4ec020aa0a59edf0452d873e26c&site=xueshu_se.
- Chancey, E. T., J. P. Bliss, Y. Yamani, and H. A. H. Handley. 2017. "Trust and the Compliance–Reliance Paradigm: The Effects of Risk, Error Bias, and Reliability on Trust and Dependence." *Human Factors* 59, no. 3: 333–345. <https://doi.org/10.1177/0018720816682648>.
- Chavaillaz, A., D. Wastell, and J. Sauer. 2016. "System Reliability, Performance and Trust in Adaptable Automation." *Applied Ergonomics* 52: 333–342.
- Cramer, H., V. Evers, S. Ramlal, et al. 2008. "The Effects of Transparency on Trust in and Acceptance of a Content-Based Art Recommender." *User Modeling and User-Adapted Interaction* 18, no. 5: 455–496. <https://doi.org/10.1007/s11257-008-9051-3>.
- Daziano, R. A., M. Sarrias, and B. Leard. 2017. "Are Consumers Willing to Pay to Let Cars Drive for Them? Analyzing Response to Autonomous Vehicles." *Transportation Research Part C: Emerging Technologies* 78: 150–164. <https://doi.org/10.1016/j.trc.2017.03.003>.
- Dixon, S. R., and C. D. Wickens. 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: 474–486.
- Dixon, S. R., C. D. Wickens, and J. S. McCarley. 2006. "How Do Automation False Alarms and Misses Affect Operator Compliance and Reliance?" *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, no. 1: 25–29.
- Dixon, S. R., C. D. Wickens, and J. S. McCarley. 2007. "On the Independence of Compliance and Reliance: Are Automation False Alarms Worse Than Misses?" *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49, no. 4: 564–572.
- Dzindolet, M. T., S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck. 2003. "The Role of Trust in Automation Reliance." *International Journal of Human-Computer Studies* 58, no. 6: 697–718.

- Fagnant, D. J., and K. Kockelman. 2015. "Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations." *Transportation Research Part A: Policy and Practice* 77: 167–181. <https://doi.org/10.1016/j.tra.2015.04.003>.
- Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner. 2007. "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39, no. 2: 175–191. <https://doi.org/10.3758/BF03193146>.
- Featherman, M., S. Jia, C. B. Califf, and N. Hajli. 2021. "The Impact of New Technologies on Consumers Beliefs: Reducing the Perceived Risks of Electric Vehicle Adoption." *Technological Forecasting and Social Change* 169: 120847. <https://doi.org/10.1016/j.techfore.2021.120847>.
- Groom, V., V. Srinivasan, C. L. Bethel, R. Murphy, L. Dole, and C. Nass. 2011. "Responses to Robot Social Roles and Social Role Framing." *2011 International Conference on Collaboration Technologies and Systems (CTS)*: 194–203.
- Guo, H., L. Song, J. Liu, et al. 2018. "Hazard-Evaluation-Oriented Moving Horizon Parallel Steering Control for Driver-Automation Collaboration During Automated Driving." *IEEE/CAA Journal of Automatica Sinica* 5, no. 6: 1062–1073. <https://doi.org/10.1109/JAS.2018.7511225>.
- Hallahan, K. 1999. "Seven Models of Framing: Implications for Public Relations." *Journal of Public Relations Research* 11, no. 3: 205–242. https://doi.org/10.1207/s1532754xjpr1103_02.
- Hancock, P. A., D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman. 2011. "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53, no. 5: 517–527.
- He, X., J. Stapel, M. Wang, and R. Happee. 2022. "Modelling Perceived Risk and Trust in Driving Automation Reacting to Merging and Braking Vehicles." *Transportation Research Part F: Traffic Psychology and Behaviour* 86: 178–195. <https://doi.org/10.1016/j.trf.2022.02.016>.
- Hergeth, S., L. Lorenz, R. Vilimek, and J. F. Krems. 2016. "Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58, no. 3: 509–519. <https://doi.org/10.1177/0018720815625744>.
- Hoff, K. A., and M. Bashir. 2015. "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, no. 3: 407–434.
- Holthausen, B. E., P. Wintersberger, B. N. Walker, and A. Riener. 2020. "Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation." *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*: 40–47. <https://doi.org/10.1145/3409120.3410637>.
- Jarosch, O., S. Paradies, D. Feiner, and K. Bengler. 2019. "Effects of Non-Driving Related Tasks in Prolonged Conditional Automated Driving – A Wizard of Oz On-Road Approach in Real Traffic Environment." *Transportation Research Part F: Traffic Psychology and Behaviour* 65: 292–305. <https://doi.org/10.1016/j.trf.2019.07.023>.
- Jian, J.-Y., A. M. Bisantz, and C. G. Drury. 2000. "Foundations for an Empirically Determined Scale of Trust in Automated Systems." *International Journal of Cognitive Ergonomics* 4, no. 1: 53–71. https://doi.org/10.1207/S15327566IJCE0401_04.
- Johnson, J. D., J. Sanchez, A. D. Fisk, and W. A. Rogers. 2004. "Type of Automation Failure: The Effects on Trust and Reliance in Automation." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 48, no. 18: 2163–2167.
- Kidd, C., and C. David. 2003. Sociable Robots: The Role of Presence and Task in Human-Robot Interaction.
- Kim, H., J. Hong, and S. Lee. 2024. "Investigating Impact of Situation Awareness-Based Displays of Semi-Autonomous Driving in Urgent Situations." *Transportation Research Part F: Traffic Psychology and Behaviour* 105: 454–472. <https://doi.org/10.1016/j.trf.2024.07.018>.
- Large, D. R., G. Burnett, D. Salanitri, A. Lawson, and E. Box. 2019. *A Longitudinal Simulator Study to Explore Drivers' Behaviour in Level 3 Automated Vehicles*. 11.
- Lee, J., and N. Moray. 1992. "Trust, Control Strategies and Allocation of Function in Human-Machine Systems." *Ergonomics* 35, no. 10: 1243–1270. doi:10/czxpww.
- Lee, J. D., and K. A. See. 2004. "Trust in Automation: Designing for Appropriate Reliance." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, no. 1: 50–80.
- Ma, R., and D. B. Kaber. 2007. "Effects of In-Vehicle Navigation Assistance and Performance on Driver Trust and Vehicle Control." *International Journal of Industrial Ergonomics* 37, no. 8: 665–673. <https://doi.org/10.1016/j.ergon.2007.04.005>.
- Ma, S., W. Zhang, Z. Yang, et al. 2021. "Take over Gradually in Conditional Automated Driving: The Effect of Two-Stage Warning Systems on Situation Awareness, Driving Stress, Takeover Performance, and Acceptance." *International Journal of Human-Computer Interaction* 37, no. 4: 352–362. <https://doi.org/10.1080/10447318.2020.1860514>.
- Madhavan, P., D. A. Wiegmann, and F. C. Lacson. 2006. "Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48, no. 2: 241–256. <https://doi.org/10.1518/00187200677724408>.
- Maltz, M., and D. Shinar. 2003. "New Alternative Methods of Analyzing Human Behavior in Cued Target Acquisition." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 45, no. 2: 281–295. <https://doi.org/10.1518/hfes.45.2.281.27239>.
- Masalonis, A. J., and R. Parasuraman. 1999. "Trust as a Construct for Evaluation of Automated Aids: Past and Future Theory and Research." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 43, no. 3: 184–187. <https://doi.org/10.1177/154193129904300312>.
- Mayer, R. C., J. H. Davis, and F. D. Schoorman. 1995. "An Integrative Model of Organizational Trust." *Academy of Management Review* 20, no. 3: 709. <https://doi.org/10.2307/258792>.
- Meyer, J. 2001. "Effects of Warning Validity and Proximity on Responses to Warnings." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 43, no. 4: 563–572. <https://doi.org/10.1518/001872001775870395>.
- Miller, D., A. Sun, and W. Ju. 2014. "Situation Awareness With Different Levels of Automation." *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*: 688–693. <https://doi.org/10.1109/SMC.2014.6973989>.
- Mishler, S., and J. Chen. 2023. "Effect of Automation Failure Type on Trust Development in Driving Automation Systems." *Applied Ergonomics* 106: 103913.
- Morris, D. M., J. M. Erno, and J. J. Pilcher. 2017. "Electrodermal Response and Automation Trust During Simulated Self-Driving Car Use." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, no. 1: 1759–1762. <https://doi.org/10.1177/1541931213601921>.
- Paepcke, S., and L. Takayama. 2010. "Judging a Bot by Its Cover: An Experiment on Expectation Setting for Personal Robots." *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*: 45–52.
- Parasuraman, R., and V. Riley. 1997. "Humans and Automation: Use, Misuse, Disuse, Abuse." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39, no. 2: 230–253.
- Parasuraman, R., T. B. Sheridan, and C. D. Wickens. 2008. "Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs." *Journal of Cognitive Engineering and Decision Making* 2, no. 2: 140–160.
- Petersen, L., L. Robert, X. J. Yang, and D. Tilbury. 2019. "Situational Awareness, Driver's Trust in Automated Driving Systems and

- Secondary Task Performance.” *SAE International Journal of Connected and Automated Vehicles* 2, no. 2: 129–141. <https://doi.org/10.4271/12-02-02-0009>.
- Radlmayr, J., C. Gold, L. Lorenz, M. Farid, and K. Bengler. 2014. “How Traffic Situations and Non-Driving Related Tasks Affect the Take-Over Quality in Highly Automated Driving.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58, no. 1: 2063–2067. <https://doi.org/10.1177/1541931214581434>.
- Rice, S., and J. S. McCarley. 2011. “Effects of Response Bias and Judgment Framing on Operator Use of an Automated Aid in a Target Detection Task.” *Journal of Experimental Psychology: Applied* 17, no. 4: 320–331. <https://doi.org/10.1037/a0024243>.
- Rovira, E., A. Cross, E. Leitch, and C. Bonaceto. 2014. “Displaying Contextual Information Reduces the Costs of Imperfect Decision Automation in Rapid Retasking of ISR Assets.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 56, no. 6: 1036–1049. <https://doi.org/10.1177/0018720813519675>.
- Ryan, G. W., and H. R. Bernard. 2003. “Techniques to Identify Themes.” *Field Methods* 15, no. 1: 85–109. <https://doi.org/10.1177/1525822X02239569>.
- Shi, X., Z. Wang, X. Li, and M. Pei. 2021. “The Effect of Ride Experience on Changing Opinions Toward Autonomous Vehicle Safety.” *Communications in Transportation Research* 1: 100003. <https://doi.org/10.1016/j.commtr.2021.100003>.
- Stapel, J., A. Gentner, and R. Happee. 2022. “On-Road Trust and Perceived Risk in Level 2 Automation.” *Transportation Research Part F: Traffic Psychology and Behaviour* 89: 355–370. <https://doi.org/10.1016/j.trf.2022.07.008>.
- Tannen, D., and C. Wallat. 1987. “Interactive Frames and Knowledge Schemas in Interaction: Examples From a Medical Examination/Interview.” *Social Psychology Quarterly* 50, no. 2: 205. <https://doi.org/10.2307/2786752>.
- Taylor, R. M. 1990. “Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design.” In *Situational Awareness*, edited by E. Salas, (1st ed., 111–128. Routledge. <https://doi.org/10.4324/9781315087924-8>.
- Tversky, A., and D. Kahneman. 1981. “The Framing of Decisions and the Psychology of Choice.” *Science* 211, no. 4481: 453–458. <https://doi.org/10.1126/science.7455683>.
- de Visser, E. J., R. Pak, and T. H. Shaw. 2018. “From ‘Automation’ to ‘Autonomy’: The Importance of Trust Repair in Human–Machine Interaction.” *Ergonomics* 61, no. 10: 1409–1427. <https://doi.org/10.1080/00140139.2018.1457725>.
- Vliegthart, R. 2012. “Framing in Mass Communication Research – An Overview and Assessment.” *Sociology Compass* 6, no. 12: 937–948. <https://doi.org/10.1111/soc4.12003>.
- Walker, F., A. Boelhouwer, T. Alkim, W. B. Verwey, and M. H. Martens. 2018. “Changes in Trust After Driving Level 2 Automated Cars.” *Journal of Advanced Transportation* 2018: 1–9. <https://doi.org/10.1155/2018/1045186>.
- Wang, L., G. A. Jamieson, and J. G. Hollands. 2009. “Trust and Reliance on an Automated Combat Identification System.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 51, no. 3: 281–291. <https://doi.org/10.1177/0018720809338842>.
- Washburn, A., A. Adeleye, T. An, and L. D. Riek. 2020. “Robot Errors in Proximate HRI: How Functionality Framing Affects Perceived Reliability and Trust.” *ACM Transactions on Human-Robot Interaction* 9, no. 3: 1–21. <https://doi.org/10.1145/3380783>.
- Xing, Y., C. Lv, D. Cao, and P. Hang. 2021. “Toward Human-Vehicle Collaboration: Review and Perspectives on Human-Centered Collaborative Automated Driving.” *Transportation Research Part C: Emerging Technologies* 128: 103199. <https://doi.org/10.1016/j.trc.2021.103199>.
- Xu, Z., K. Zhang, H. Min, Z. Wang, X. Zhao, and P. Liu. 2018. “What Drives People to Accept Automated Vehicles? Findings From a Field Experiment.” *Transportation Research Part C: Emerging Technologies* 95: 320–334. <https://doi.org/10.1016/j.trc.2018.07.024>.
- Yan, X., X. Li, Y. Liu, and J. Zhao. 2014. “Effects of Foggy Conditions on Drivers’ Speed Control Behaviors At Different Risk Levels.” *Safety Science* 68: 275–287. <https://doi.org/10.1016/j.ssci.2014.04.013>.
- Zhang, C., R. Tao, H. Zhao, et al. 2022. “Two Inconsistent Rounds of Feedback Enhance the Framing Effect: Coding Two Consecutive Outcome Evaluations.” *International Journal of Psychophysiology* 182: 47–56. <https://doi.org/10.1016/j.ijpsycho.2022.09.012>.
- Zhang, J., and Y. Wu. 2005. “Likelihood-Ratio Tests for Normality.” *Computational Statistics & Data Analysis* 49, no. 3: 709–721. <https://doi.org/10.1016/j.csda.2004.05.034>.
- Zhou, Y., X. Cui, W. Qu, and Y. Ge. 2022. “The Effect of Automation Trust Tendency, System Reliability and Feedback on Users’ Phishing Detection.” *Applied Ergonomics* 102: 103754. <https://doi.org/10.1016/j.apergo.2022.103754>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.