When Three Share the Ride: Social Meaning-Making and Real-Time Trust-Loss Signals in Automated Vehicles

Abstract

Trust is a critical determinant of public acceptance of automated vehicles, particularly in situations involving uncertainty, elevated risk, or time pressure. Although much of the existing research conceptualizes trust as an individual cognitive process, the growing emergence of shared autonomous mobility highlights the need to understand how trust is constructed within groups. This study examines how situational urgency, control modality, and social interaction influence trust during shared rides in an automated vehicle. Thirty-six participants completed a within-subjects virtual reality simulation in groups of three, experiencing both automated and manual driving under low- and high-urgency scenarios. Trust was assessed through real-time button-press behavior, post-trial questionnaires, and thematic analysis of group conversations. Results showed that trust was lowest during automated low-urgency conditions and highest during manual high-urgency conditions, with button-press behavior sensitively capturing moment-to-moment loss of trust. Qualitative findings indicated that trust was co-constructed through group discussions that evaluated vehicle behavior, situational interpretation, and perceived transparency of the system. Overall, the findings demonstrate that trust in automated vehicles is both context-dependent and socially embedded. Participants displayed reduced trust in automation when situational risk was low and greater trust in manual control when urgency was high, challenging assumptions that time pressure universally promotes reliance on automation. These insights highlight the need for automated vehicles that adapt to behavioral and interpersonal trust cues in multi-occupant scenarios.

Keywords: trust, automated vehicles, group dynamics, human–robot interaction, virtual reality

1. Introduction

The emergence of automated vehicles (Automated vehicles (AVs)) marks a significant development in contemporary transportation systems, with the potential to improve road safety and enhance mobility efficiency. A substantial proportion of traffic crashes are attributed to human error, with estimates exceeding 94% [1, 2, 3]. By reducing human involvement in complex driving tasks, AVs are expected to enable smoother traffic flow [4], greater compliance with traffic regulations [5], and increased in-vehicle productivity through support for non-driving-related tasks (non-driving-related tasks (NDRTs)) [6, 7, 8]. Despite these anticipated benefits, public trust in AVs remains limited [9], particularly in situations characterized by uncertainty or elevated risk [10].

Existing research on trust in AVs has predominantly examined individual passengers' judgments and behavioral responses to automation [11, 12, 13]. However, as shared automated vehicles (shared automated vehicles (SAVs)) become increasingly plausible, understanding how trust is shaped within groups of co-present passengers is gaining importance. Shared rides create opportunities for social reassurance, verbal negotiation, and collective interpretation of driving events. These interpersonal processes may substantially influence trust formation, yet they remain underexplored in the context of automated mobility [14]. Furthermore, prior work has not combined group-level conversational analysis with real-time behavioral markers of trust during simulated SAV journeys involving triads of passengers. Addressing this gap is essential for conceptualizing trust in AVs as a socially embedded, context-sensitive phenomenon rather than a purely individual judgment.

The present study investigates how situational urgency, control modality, and interpersonal interaction shape trust during shared rides in an AV. Thirty-six participants completed a within-subjects virtual reality (virtual reality (VR)) driving simulation in groups of three, experiencing automated and manual driving conditions under low- and high-urgency scenarios. Trust was assessed using three complementary approaches: (1) real-time button-press behavior indicating moment-to-moment changes in perceived safety, (2) post-trial questionnaires capturing reflective evaluations, and (3) thematic analysis of group conversations. Together, these methods offer an integrative perspective on how passengers experience and express trust during shared automated mobility.

2. Background

2.1. Trust in Automated Vehicles

Trust in automation is shaped by dispositional, situational, and learned factors [11, 15, 16]. In the context of AVs, studies have examined how perceived reliability, predictability, transparency, and user experience influence the willingness to rely on automated systems [12, 16]. Trust is commonly conceptualized as the belief that an automated system will act in ways that support user goals under uncertainty [15]. It is understood as dynamic and subject to recalibration based on system performance [11, 17]. Achieving calibrated trust, neither overreliance nor unwarranted distrust, is critical for safe and effective interaction with AVs [16, 18].

2.2. Measuring Trust: From Self-Report to Real-Time Indicators

Trust in AVs has traditionally been assessed using questionnaires [12, 19, 20, 21], which provide valuable but retrospective insight. Increasingly, researchers highlight the need for real-time behavioral measures of trust [22, 23]. Gaze behavior has been widely explored, with higher trust often associated with reduced road monitoring and increased engagement in NDRTs [24, 25, 26], and lower trust linked to vigilant scanning and physiological stress [27, 25]. Physiological signals, such as electrodermal activity (EDA), respiration, and heart rate, have also been used to infer trust and user state [28]. However, gaze–trust correlations remain inconsistent [29], leaving the need for additional trust-sensitive behavioral indicators.

The present study incorporates a real-time button-press mechanism, allowing participants to signal moments of reduced trust during driving. This simple but unobtrusive input offers a promising complement to questionnaire-based assessments, enabling fine-grained observation of trust fluctuations as they occur.

2.3. Trust in Context: Situational Urgency

Trust in AVs is shaped not only by system performance but also by contextual factors [11]. In low-urgency scenarios, consistent and predictable behavior can reinforce trust over time [30]. High-urgency situations, including sudden hazards, unclear feedback, or system hesitation, pose greater challenges, potentially amplifying stress and eroding confidence [31, 32]. Perceptions of safety depend on the assertiveness, smoothness, and transparency of AV responses [33, 34, 35]. Although trust may be restored following corrective action [36, 37, 38], empirical understanding of how trust evolves during and after high-urgency events remains limited.

2.4. Trust as a Social Process

Shared mobility settings introduce interpersonal dynamics that shape passenger perception. Social proof suggests that people look to others' reactions for guidance during uncertainty [39]. In SAVs, conversational cues, such as explicit discussion, reassurance, or expressions of doubt, can influence trust judgments [14]. Passengers may respond not only to the vehicle but to each other's interpretations, comfort levels, and behavioral signals [40, 41]. In some cases, anxiety about co-passengers can outweigh concerns about the vehicle itself [41]. We conceptualize these processes as socially embedded trust: an emergent, dynamic phenomenon shaped through real-time interpersonal coordination.

2.5. Aim of the Study

This study examines trust in AVs as a dynamic, context-sensitive, and socially embedded process. Moving beyond individual assessments, it investigates how trust is experienced and expressed in shared rides and how it fluctuates in response to situational urgency, control modality, and interpersonal interaction. The following research questions guided the study:

1. How does trust in AVs fluctuate in response to situational urgency and control modality?
2. Can button-press behavior function as a real-time behavioral indicator of trust loss?
3. How is trust in AV performance co-constructed through group conversation during shared rides?

3. Method

3.1. Participants

A power analysis using G*Power 3 [42] indicated that 36 participants were required to achieve 95% power for a repeated-measures design with a medium effect size ($f = 0.25$, $d = 0.5$, $\alpha = .05$). Participants (ages 18–30) were recruited over 12.5 weeks via mixed convenience sampling. Eligibility criteria included holding a valid driver's license, English proficiency, normal or corrected-to-normal vision, and no prior experience as a passenger in an automated vehicle. Participants received two research credits as compensation. Some enrolled individually, while others participated with friends or acquaintances. Ethical approval for the study was obtained from the Psychology Research Ethics Committee of the [anonymous university].

3.2. Apparatus

The experimental setup consisted of three Next Level Racing seats equipped with Logitech driving hardware and Varjo headsets with integrated eye tracking. The simulation was rendered in high-fidelity virtual reality (VR) using Unity and the Coupled Simulator, an open-source platform designed to support multi-user research in traffic environments [43]. Button-press input was collected via handheld VR controllers. Group conversations were captured using a Blue Snowball microphone and recorded with Open Broadcaster Software (open broadcaster software studio (OBS)).

3.3. Materials and Stimuli

The driving environment consisted of a realistic urban–highway setting presented in VR. Situational urgency was manipulated through route-based events. In the low-urgency scenario, participants encountered a routine exit from a highway back into the city (Figure 1). In the high-urgency scenario, they encountered two critical events: a roadblock requiring an evasive maneuver and a flipped car obstructing part of the lane (Figure 2). These events were selected to establish a clear contrast in perceived risk and time pressure.

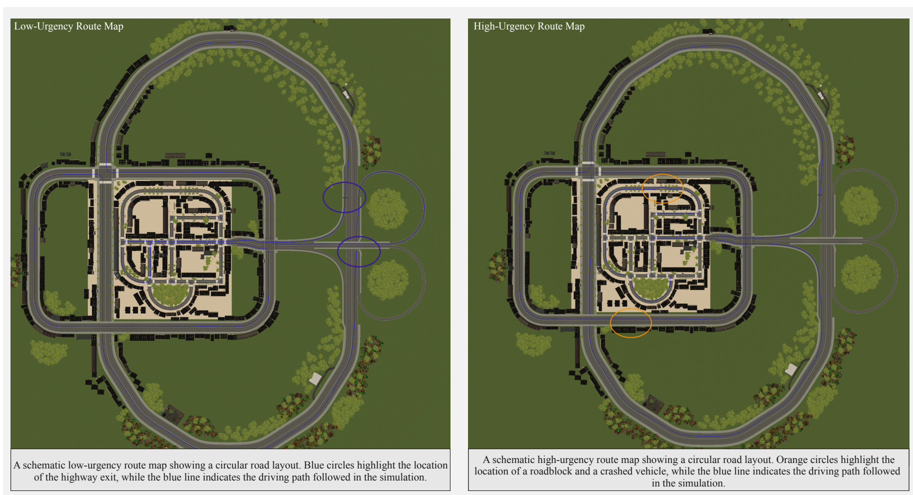

Figure 1: Schematic route maps for the low- and high-urgency scenarios. Blue circles indicate highway exits (low urgency). Orange circles indicate roadblocks and a flipped car (high urgency).

Note. The maps illustrate how situational urgency was manipulated using specific road events distributed along fixed routes.

Figure 2: Screenshot examples of low- and high-urgency events: (A) Exit ramp (low urgency), (B) roadblock (high urgency), and (C) flipped car (high urgency).

Note. Images illustrate key moments used to manipulate situational urgency.

3.4. Measures

Button-press behavior. Participants used the right-hand VR controller to press and hold a trigger whenever they experienced a loss of trust, releasing the button when trust was restored. Button-press time series were aggregated per trial and visualized across urgency and control conditions to characterize moment-to-moment trust dynamics.

Trust questionnaire. Trust was assessed using an adapted 7-item Trust in Automation Scale [44, 45]. Participants completed a baseline questionnaire before the first trial and a post-trial questionnaire after each scenario. Items (e.g., "I trusted the automated car/driver") were rated on a 7-point Likert scale. In manual trials, only the two passengers completed the post-trial questionnaire to avoid driver self-assessment bias.

Group conversation. After each trial, participants engaged in brief group discussions reflecting on the driving episode and their trust in the system. Conversations were recorded and transcribed orthographically using Otter.ai. Transcriptions preserved speech features such as hesitations, false starts, and emphasis to support subsequent qualitative analysis.

3.5. Design

The study employed a 2 × 2 within-subjects design with factors of driving mode (manual vs. automated) and situational urgency (low vs. high). Each triad completed four 10-minute trials (Manual–Low, Manual–High, Automated–Low, Automated–High). Trial order and seat assignments (driver, front passenger, rear passenger) were randomized.

Figure 3: Lab setup used in the study, showing VR headsets, seating arrangement, and driving hardware used during the simulation.

3.6. Procedure

Upon arrival, participants provided informed consent and completed a baseline trust questionnaire. After a brief orientation to the equipment, they completed a practice trial to familiarize themselves with the VR environment and button-press mechanism.

During manual trials, one participant controlled the vehicle using a steering wheel and pedals, while the two passengers used VR controllers to report trust loss. During automated trials, the vehicle operated autonomously and all three participants could respond using the button-press mechanism. Participants were instructed to converse naturally during the trials and to press the trigger whenever they experienced a loss of trust in the system or driver.

Although English proficiency was an inclusion criterion, participants were encouraged to communicate naturally. As all were based in the Netherlands, some triads occasionally switched between English and Dutch; this was not restricted. After each trial, participants completed the post-trial trust questionnaire. A break was provided mid-session.

The full experimental timeline is shown in Figure 4.

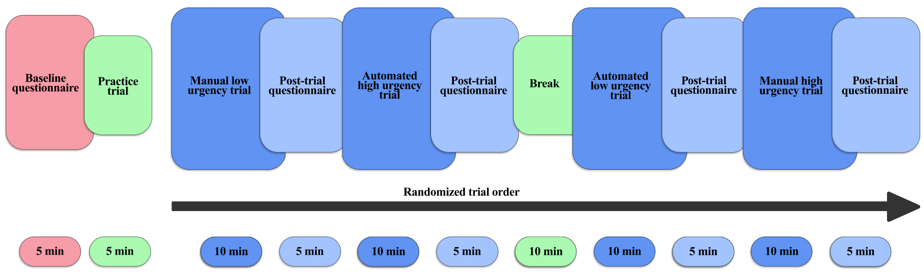

Figure 4: Timeline of study activities. Participants completed a baseline questionnaire, practice trial, four test trials, and a post-trial questionnaire after each scenario. A break occurred after the second trial.

Note. Trial order and seat assignments were randomized across triads.

3.7. Analysis

A multi-method analytical approach was used to examine trust dynamics. Questionnaire data were analyzed using repeated-measures ANOVAs to assess the effects of situational urgency and control modality on trust scores. Paired-samples t-tests were conducted for post-hoc comparisons, and an additional ANOVA including the baseline condition was performed to assess trust changes in high-urgency scenarios. Internal consistency of the scales was evaluated using Cronbach's alpha.

Button-press behavior was analyzed using repeated-measures MANOVA across the four trial types. Time-locked visualizations illustrated button-press activity in relation to key scenario events. Pearson correlations were calculated between post-trial trust scores and the percentage of trial time in which participants pressed the trust-loss button to assess convergence between subjective and behavioral measures.

Qualitative data were analyzed using reflexive thematic analysis following the six-phase framework proposed by Braun and Clarke [46]. Initial codes were generated at both semantic and latent levels and clustered into candidate themes. Themes were reviewed for internal coherence and relevance, then refined and clearly defined. Two researchers each coded half of the transcripts. Dutch-language transcripts were coded by the Dutch-speaking researcher. To enhance reflexivity and consistency, two transcripts from each set were randomly selected for cross-coding. Final themes were developed collaboratively to provide a coherent account addressing the research questions.

4. Results

Results are reported across three domains: (1) trust questionnaire scores, (2) button-press behavior as a real-time indicator of trust loss, and (3) themes identified in group conversations.

4.1. Trust Across Urgency and Control Conditions

Internal consistency of the trust questionnaire was high (baseline: $\alpha = .93$; post-trial: $\alpha = .94$). A paired-samples t-test indicated that trust was significantly higher following high-urgency trials ($M = 4.41$, $SD = 0.97$) compared with low-urgency trials ($M = 4.03$, $SD = 0.97$), $t(35) = 2.39$, $p = .022$, $d = 0.40$.

To examine combined effects of urgency and control modality, a repeated-measures ANOVA compared four conditions: manual high urgency ($M = 4.67$, $SD = 1.09$), manual low urgency ($M = 4.54$, $SD = 1.23$), automated high urgency ($M = 4.10$, $SD = 1.23$), and automated low urgency ($M = 3.71$, $SD = 1.31$). Sphericity violations were corrected ($\epsilon = 0.669$). A significant main effect of condition was observed, $F(2.01, 46.17) = 4.80$, $p = .013$, $\eta_p^2 = .17$. Bonferroni-adjusted contrasts showed that trust in manual high-urgency trials was significantly higher than in automated low-urgency trials ($\Delta M = 0.96$, $p = .024$). No other pairwise differences reached significance. Figure 5 illustrates these patterns.

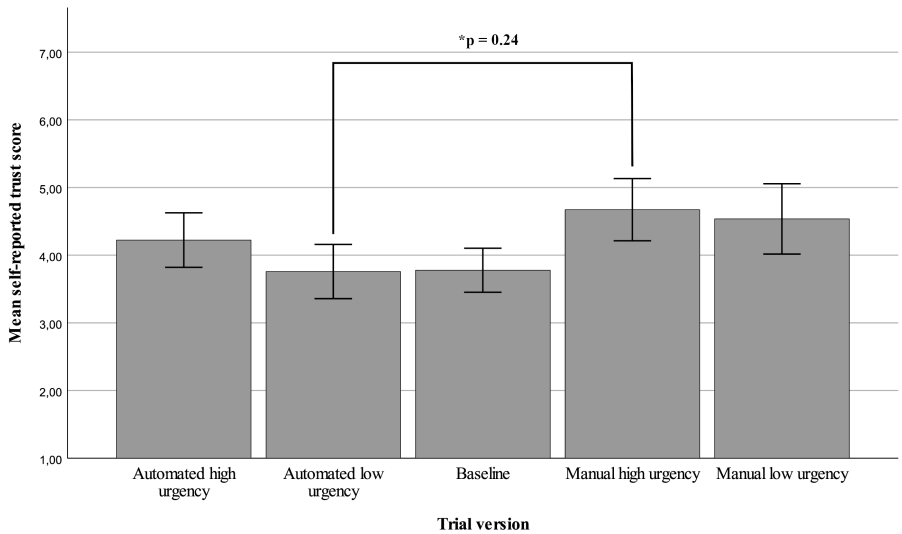

Figure 5: Trust questionnaire scores across urgency and control conditions. Error bars represent standard error.

To examine trust changes relative to participants' initial expectations, we conducted a follow-up ANOVA including baseline, automated high urgency, and manual high urgency. Results showed a significant effect of condition, $F(2, 46) = 5.82$, $p = .006$, $\eta_p^2 = .20$. Trust increased from baseline ($M = 3.77$, $SE = 0.21$) to automated high urgency ($M = 4.05$, $SE = 0.25$), and was highest in manual high urgency ($M = 4.67$, $SE = 0.22$; see Figure 6). The baseline–manual comparison was significant ($p = .005$, $d = 0.72$); other contrasts were not ($p > .19$).

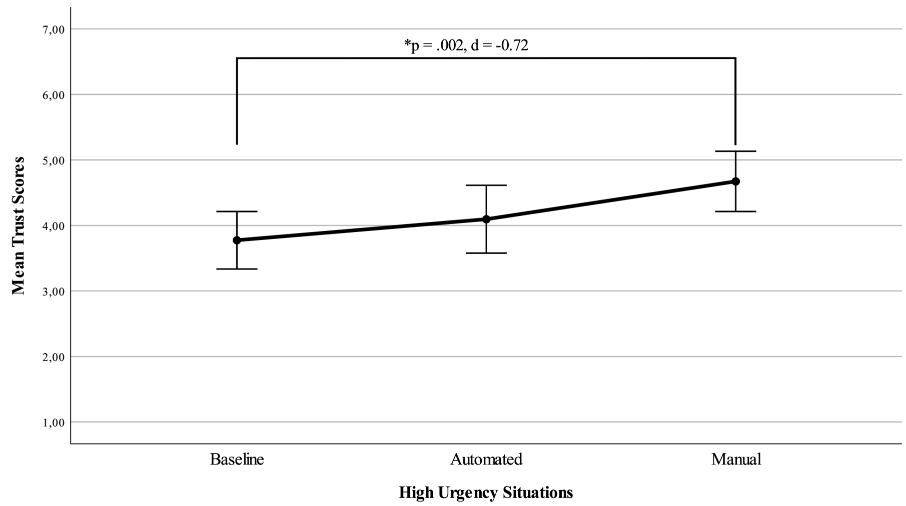

Figure 6: Estimated marginal means of trust across baseline, automated, and manual conditions in high-urgency scenarios (95% CIs shown).

Because manual post-trial questionnaires were completed only by passengers ($N = 24$), additional paired t-tests were conducted. Trust increased significantly from baseline to manual high urgency, $t(23) = -3.54$, $p = .002$. Trust also tended to increase from baseline to automated high urgency, although this difference approached significance only at trend level, $t(35) = -2.01$, $p = .052$.

Collapsing across urgency, trust was significantly higher in manual trials ($M = 4.60$, $SD = 1.00$) compared with automated trials ($M = 3.90$, $SD = 1.20$), $t(23) = 2.52$, $p = .019$, $d = 0.51$. Taken together, these results indicate that trust was highest in manual high-urgency scenarios and lowest in automated low-urgency scenarios, suggesting a stronger effect of control modality than urgency alone.

4.2. Button-Press Behavior

Button-press activity captured moment-to-moment trust fluctuations. On average, participants held the trust-loss button for 2.70% of total trial time. Condition means were as follows: automated low urgency ($M = 5.57\%$, $SD =$

5.57), automated high urgency ($M = 2.51\%$, $SD = 2.81$), manual low urgency ($M = 1.55\%$, $SD = 2.46$), and manual high urgency ($M = 1.16\%$, $SD = 1.85$).

A repeated-measures MANOVA revealed a significant multivariate effect of condition, Pillai's Trace = 0.398, $F(3, 33) = 7.28$, $p < .001$, $\eta_p^2 = .398$. Greenhouse–Geisser corrected univariate tests (due to sphericity violations: $\chi^2(5) = 57.91$, $p < .001$) showed a significant effect of trial type on button-press duration, $F(1.47, 48.53) = 15.74$, $p < .001$, $\eta_p^2 = .310$.

Bonferroni-corrected comparisons revealed that automated low-urgency trials elicited significantly more button-press behavior than all other conditions (all $p < .01$). No other contrasts were significant. Figure 7 shows time-locked peaks aligned with scenario events.

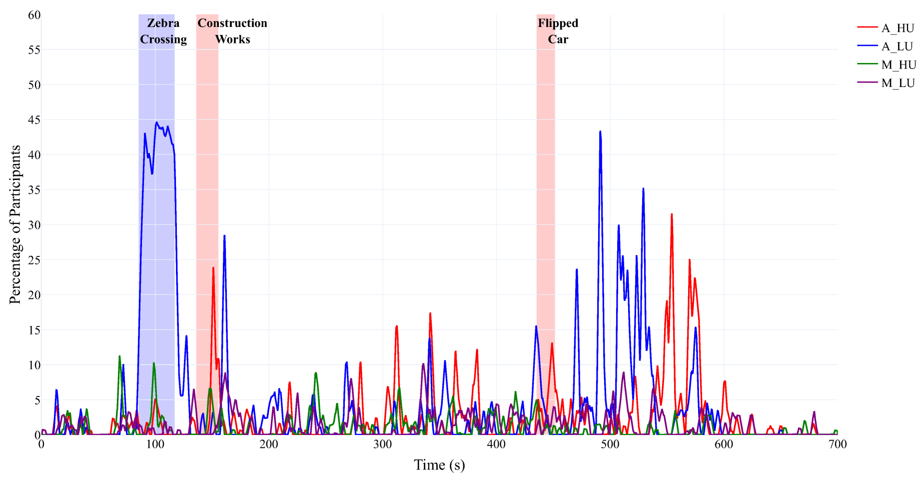

Figure 7: Percentage of participants holding the trust-loss button over time across trial conditions. Red areas indicate high-urgency events.

Overall, button-press behavior was more frequent in automated than manual conditions, with the automated low-urgency scenario generating the highest levels of real-time trust loss.

4.3. Convergence Between Questionnaire and Behavioral Measures

To assess convergence between subjective and behavioral indicators of trust, correlations were computed between post-trial questionnaire scores and button-press duration. Only automated trials were included, as they provided button-press data for all participants.

A significant negative correlation emerged, $r(118) = -0.41$, $p < .001$, indicating that lower reported trust was associated with longer button-press durations. Figure 8 visualizes this negative relationship.

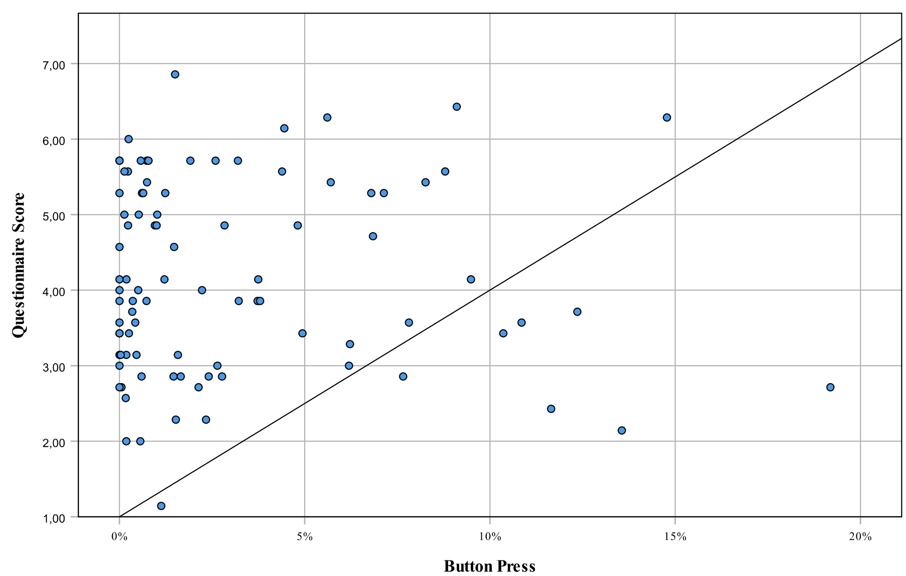

Figure 8: Correlation between button-press duration and trust scores in automated trials. Each point represents one participant–trial.

These results support button-press behavior as a valid real-time indicator of trust loss in automated driving contexts.

4.4. Group Conversation Themes

Thematic analysis of post-trial discussions revealed several recurrent themes characterizing group-level meaning-making processes. Table 1 summarizes the themes, descriptions, and prevalence across conversations.

Table 1: Themes Identified in Group Conversations

Theme	**Description**	**#C**	**#G**	**%G**
Navigation uncertainty	Comments about route choice and navigation logic	15	8	66
Automated vehicle behavior	Shared evaluations of AV performance (speed, lane keeping, maneuvers)	37	11	91
Environmental factors	References to missing traffic, pedestrians, or ambient realism	8	6	50
Relating to real life	Comparisons to participants' everyday driving experiences	6	6	50
Vehicle feedback	Observations about missing or unclear system cues (e.g., blinkers)	18	8	66
Comparing manual vs. automated	Direct comparisons between perceived control and AV operation	9	7	58

Note. #C = number of conversations; #G = number of groups; %G = percentage of groups referencing theme.

5. Discussion

This study examined how trust in automated vehicles (AVs) is shaped by situational urgency, control modality, and social dynamics during shared rides. By combining real-time button-press input, post-trial questionnaires, and thematic analysis of group conversations, we captured both individual and collective expressions of trust as they unfolded in a shared automated mobility context.

5.1. Urgency, Control Modality, and Perceived Agency

Addressing RQ1, we investigated how trust fluctuated across urgency levels and control conditions. Contrary to initial expectations that high-urgency scenarios would erode trust, trust was higher in high-urgency than in low-urgency trials, particularly when a human was in control. Across measures, participants reported consistently higher trust in manual conditions, suggesting that perceived human agency plays a central role in trust formation.

Even though only one group member actively controlled the vehicle, the presence of a human driver appeared to reassure the other passengers. Observing human-driven behavior seemed sufficient to enhance perceived reliability, indicating that trust was influenced by how control was attributed rather than by direct control experience. This aligns with prior work arguing that trust is shaped not simply by performance, but by how responsibility is distributed between humans and automation [11, 15, 47].

Notably, trust in manual high-urgency trials exceeded baseline levels, suggesting that urgency can amplify trust when participants see a human agent managing the situation. By contrast, trust in automated high-urgency trials did not differ significantly from baseline. Urgency alone therefore did not undermine trust, but it also did not substantially deepen it in the absence of perceived human agency. These findings suggest that participants evaluated conditions primarily through a control lens: what mattered was not only what the system did, but who (or what) was perceived to be "in charge".

Group conversations reinforced these perceptions. Participants frequently compared system and manual responses (e.g., "we would have braked sooner"), making control differences highly salient. Low-urgency automated contexts, in particular, seemed to invite skepticism: automation was sometimes perceived as unnecessary or intrusive when the environment did not demand rapid action. Overall, the data suggest that urgency effects on trust are best understood as a function of perceived control and social meaning-making, rather than situational demand alone.

5.2. Button Presses as Real-Time Indicators of Trust

Addressing RQ2, we evaluated whether button-press behavior functioned as a real-time indicator of trust loss. The highest levels of button-press activity occurred in automated low-urgency trials, where participants likely had greater cognitive capacity to monitor the vehicle's behavior and question its decisions. This condition also showed the greatest variability in button-press duration, suggesting that some participants reacted strongly while others did not press at all. Such variability is consistent with models of cognition-based trust [11], which emphasize individual differences in how users evaluate automation under low-stress conditions.

Button-press patterns converged with questionnaire data, particularly in automated trials, where a significant negative correlation between button-press duration and post-trial trust scores indicated that longer button holding was associated with lower reported trust. This supports the interpretation of button presses as a meaningful behavioral index of trust loss, rather than a noisy or purely exploratory signal.

However, the qualitative data highlight that button presses were embedded in social practices. Participants occasionally discussed whether or when to press, effectively treating the mechanism as a shared monitoring tool rather than a strictly individual signal. Seating position may also have played a role: front-seat passengers, with more direct visual access to the road, appeared particularly engaged in monitoring and signaling. Conversely, the absence of button presses did not always indicate trust; in some instances, participants expressed resignation or disengagement, implying that non-pressing can also reflect withdrawal rather than confidence. Together, these findings suggest that button presses capture a valuable but context-dependent and socially embedded slice of real-time trust.

5.3. Social Dynamics and Shared Trust Formation

To address RQ3, we examined how trust in AV performance was co-constructed through group interaction. Thematic analysis identified six recurrent categories of talk that captured how passengers collaboratively made sense of the ride: navigation uncertainty, automated vehicle behavior, environmental factors, relating experiences to real life, vehicle feedback, and comparisons between automated and manual modes. These findings illustrate that passengers did not form trust judgments in isolation. Instead, they jointly interpreted events, negotiated meaning, and aligned their responses, creating a shared trust trajectory throughout each session.

Quotes follow the convention: G = group number, T = trial type (TA = automated, TM = manual), and seat number (0 = driver, 1 = front passenger, 2 = rear passenger). Time stamps (mm:ss) indicate when the statement occurred in the trial. Representative excerpts illustrating the six themes are summarized in Figure 9.

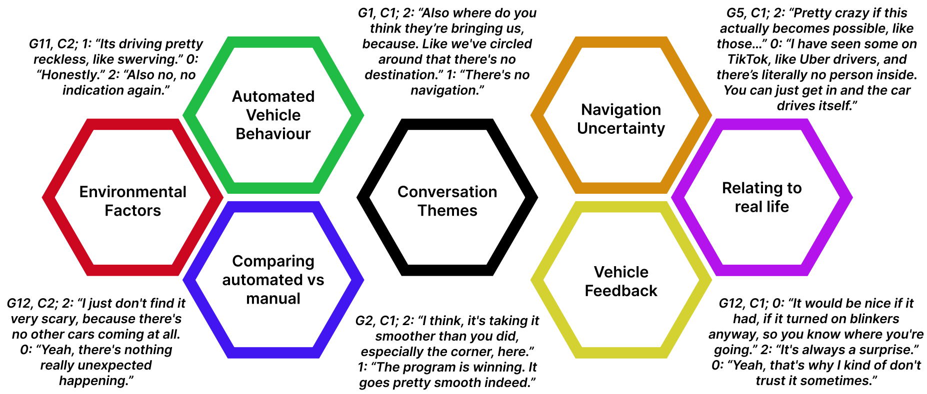

Figure 9: Representative participant quotes illustrating key aspects of trust co-construction across the six qualitative themes.

Note. Quotes follow the coding scheme: G = group number, T = trial type, and seat number (0 = driver, 1 = front passenger, 2 = rear passenger).

Navigation Uncertainty

Across eight groups (66%), passengers expressed uncertainty about the route, destination, or directional choices. These conversations frequently emerged when the vehicle appeared to circle the same area or when the next turn was ambiguous. For example, one participant remarked, *"Like we've circled around, there's no destination"* (G1, TA, 0, 10:37), followed by another noting, *"There's no navigation"* (G1, TA, 1, 10:40). Similar confusion appeared in manual trials, as when a driver asked, *"Where am I supposed to go here, left or right?"* (G2, TM, 0, 20:20). These exchanges exemplify collective sense-making under uncertainty and highlight how ambiguity can spread socially, amplifying group-level anxiety in line with collective risk perception [14, 48]. From an affective trust perspective [15], ambiguous routing heightened emotional unease, undermining confidence in the system.

Automated Vehicle Behavior

Nearly all groups (91%) scrutinized the AV's driving behavior, commenting on lane changes, speed variation, stopping logic, and the absence of signals. For example, in one triad participants exclaimed, *"Oh, we are going straight. Wait,* *what the heck?"* (G1, TA, 2, 11:10), followed by *"Still no turning signal"* (G1, TA, 0, 11:19). Concerns about unsafe or unclear stopping also surfaced: *"It's not the best place to stop"* (G5, TA, 2, 49:27). However, not all evaluations were negative; some acknowledged smoothness, such as *"It's driving very calm"* (G2, TA, 0, 3:30). These judgments reflect analytical trust processing [15], as participants compared the system's actions to implicit rules of "good driving". Behavioral inconsistencies, however, invited group-level critique, consistent with findings that opaque or non-normative automation behavior erodes trust [49].

Environmental Factors

Half the groups (50%) commented on environmental realism, especially the absence of other vehicles or pedestrians. This absence was sometimes experienced as eerie or unrealistic. For instance, participants in G4 questioned, *"Why are there no other cars?"* (G4, TA, 2, 2:27), followed by *"It's creepy"* (G4, TA, 0, 2:30). Others discussed how emptiness reduced perceived risk: *"If there's no other cars around, I'm not really afraid"* (G5, TA, 2, 7:24). These reactions underscore the importance of ecological validity in shaping trust calibration [50]. Simplified environments limited participants' ability to assess decision-making complexity, thereby reducing trust or producing discomfort.

Relating to Real Life

Six groups (50%) contextualized the simulation by referencing real-world experiences, cultural driving norms, or media portrayals of AVs. For instance, one participant asked, *"If you have an automated car, is it legal to drink and drive?"* (G2, TA, 2, 11:02), while another envisioned future interiors: *"Self-driving car could be like four chairs facing each other"* (G3, TA, 0, 56:28). Others mentioned videos of driverless taxis: *"I have seen some on TikTok ... there's literally no person inside"* (G4, TA, 03:27). These analogical reasoning processes [15] demonstrate how participants used prior experiences and cultural narratives to interpret system behavior, highlighting how trust evolves through both direct interaction and indirect knowledge [51].

Vehicle Feedback

Eight groups (66%) emphasized the centrality of feedback—particularly signaling and anticipatory cues—for maintaining trust. Missing blinkers were repeatedly reported: *"Still no turn signals"* (G1, TA, 2, 8:39), followed by *"Seems like a major oversight"* (G1, TA, 2, 8:57). Unexplained stops further heightened uncertainty: *"What's going on? Why did we stop?"* (G2, TA, 2, 4:07). These concerns align with literature on transparency and predictability in automated systems [52]. The group-level nature of these reactions suggests that insufficient feedback undermines trust collectively—when one participant voiced doubt, others quickly reinforced or expanded the critique.

Comparing Automated vs. Manual

Seven groups (58%) explicitly compared manual and automated driving. Manual driving was sometimes perceived as more competent or predictable: *"Much better"* (G1, TM, 1, 26:21). Yet, participants also noted advantages of automation: *"I think it's taking it smoother than you did"* (G1, TA, 2, 22:00). Comparisons often became humorous or evaluative exchanges, such as noting the human driver's lapses: *"You weren't looking at all the blind spots"* (G2, TM, 2, 33:05). These dialogues illustrate analogical trust processes [15], showing how participants used human driving as a yardstick to benchmark the AV, resulting in nuanced, negotiated trust calibration within the group.

Overall, the qualitative findings position trust in AVs as a socially embedded process. Group members used conversation to interpret behavior, contextualize risk, and align responses, with individual perceptions and system cues feeding into shared meaning-making. Social dynamics thus played a key role in shaping real-time trust formation during automated and manual rides.

6. Implications for Design

The findings indicate that trust in AVs is context-sensitive, socially constructed, and partially accessible through behavioral signals. These insights suggest several implications for the design of shared automated vehicles (SAVs), particularly in situations where urgency, ambiguity, or low feedback may challenge trust.

6.1. Support Real-Time Trust Expression

The button-press mechanism provided a simple, intuitive way for passengers to indicate loss of trust in real time, capturing fine-grained responses to specific events. Time-locked peaks around critical moments demonstrate the potential of such mechanisms for monitoring passenger state. Future AV systems could incorporate unobtrusive channels—such as haptic, touch-based, or gesture-based inputs—that allow passengers to express discomfort without disrupting ongoing interaction. These inputs could serve as both diagnostic tools for designers and triggers for adaptive system responses (e.g., modulation of speed or additional explanations).

6.2. Design for Social Meaning-Making

Passengers frequently relied on each other to interpret AV behavior, particularly in ambiguous moments. Interfaces should therefore support shared understanding, not just individual comprehension. Ambient feedback such as clear signaling, route visualizations, and brief verbal or visual explanations of upcoming maneuvers can provide common ground for group discussion and reduce unnecessary speculation. Designing displays and feedback visible to all occupants, rather than only the driver's position, may be especially important in SAVs.

6.3. Enhance Contextual Responsiveness

Trust was consistently lowest in automated low-urgency scenarios, suggesting that perceived redundancy, unexplained pauses, or conservative maneuvers may be interpreted as poor performance when the environment appears benign. AVs may need to adapt their communication strategies to context: in low-event-density situations, where the vehicle's actions may seem opaque or unnecessary, additional explanatory feedback (e.g., "Slowing due to speed limit change ahead") could help maintain trust. In higher-urgency situations, timely and appropriately calibrated feedback may support the sense that the system is attentive and in control.

6.4. Account for Group Settings and Multimodal Input

Because shared rides involve multiple occupants co-constructing trust, future SAVs should be able to detect and respond to group-level dynamics. Combining multimodal indicators—such as aggregated button presses, gaze patterns, or vocal cues—may enable the system to infer when a group is collectively uneasy or divided. Adaptive transparency strategies (e.g., offering more detailed explanations when uncertainty appears high) could help stabilize trust in these moments. Designing features that acknowledge the group, rather than addressing only an individual, may further support shared trust.

6.5. Strengthen Ecological Validity in Design and Testing

Participants' reactions to the absence of other traffic and pedestrians underline the importance of ecological validity in shaping trust. For real-world deployment, AV design should align behavioral and interface cues with passengers' expectations of a rich, populated environment. Even in simulation and testing phases, including more realistic traffic actors, ambient sound, and standard safety cues (e.g., seatbelt indicators, audible alerts) can support more natural calibration of trust and make findings more transferable to real-world settings.

7. Limitations and Future Work

Several limitations should be considered when interpreting the findings of this study.

First, the coordinates and movements of the simulated vehicle were not logged. As a result, participants' real-time trust-loss signals (e.g., button presses) could not be aligned precisely with specific environmental events. This limits the ability to conduct event-based analyzes of trust dynamics. Future research should incorporate synchronized trajectory logging and time-stamped event markers to link behavioral responses directly to road situations, such as lane changes, obstacles, or unexpected stops.

Second, the measurement of real-time trust was asymmetric across conditions. In manual trials, only passengers could press the trust-loss button, whereas drivers could not signal reduced trust while actively operating the vehicle. This limits comparability between control modalities and may underestimate moments of low trust in manual driving. Subsequent studies should employ more balanced multimodal sensing strategies (e.g., haptics, physiological indicators, or verbal markers) that capture trust-related responses from all occupants, including drivers.

Third, the simulation lacked certain elements of ecological realism. The absence of other traffic participants, missing safety features (such as visible seatbelts), inconsistent audio, and occasional interface constraints may have influenced how participants calibrated their trust. These factors could have attenuated perceived risk or made the environment feel artificial or "empty". Increasing environmental fidelity, by adding surrounding traffic, pedestrians, ambient sound, and standard in-vehicle safety cues, and ensuring consistent exposure to urgency events would improve generalizability and support more robust comparisons across conditions.

Finally, conversational engagement varied substantially between groups. Some triads engaged in rich, continuous discussion, whereas others remained relatively quiet. These differences likely reflected variation in group cohesion, familiarity, comfort in VR, and individual communication styles, and they may have influenced how trust was socially constructed. Future work should systematically investigate how group composition and interaction quality shape shared trust formation, for instance by using linguistic style matching, conversation metrics, or multimodal behavioral cues.

Taken together, these limitations underscore that trust in AVs is shaped not only by situational and technological factors, but also by methodological design choices and group dynamics. Future studies should build on these insights using higher-fidelity environments, balanced measurement strategies, and more systematic manipulation of group characteristics to better understand how trust develops in shared automated mobility.

8. Conclusion

This study investigated how trust in automated vehicles (AVs) develops during shared rides by examining the combined effects of situational urgency, control modality, and social interaction. Across measures, participants trusted manual driving more than automated driving, and trust was lowest in automated low-urgency scenarios where system behavior appeared less necessary, less transparent, or insufficiently responsive to context. High-urgency situations increased trust primarily when a human was perceived to be in control, underscoring the importance of agency and accountability in trust formation.

Real-time button-press behavior provided a meaningful and sensitive indicator of moment-to-moment trust loss, particularly in automated trials. Its convergence with questionnaire scores demonstrates the value of incorporating behavioral indicators alongside traditional self-report measures when studying trust dynamics in AVs. At the same time, the qualitative analysis revealed that trust was constructed through group-level sense-making: passengers discussed vehicle behavior, interpreted ambiguous events together, and used comparisons to real-world driving to evaluate system reliability. These social processes shaped trust as much as, if not more than, the driving events themselves.

Overall, the findings highlight that trust in AVs is not solely an individual psychological judgment, but a socially embedded and context-dependent process. Designing automated vehicles and shared mobility systems will require approaches that recognize the interpersonal nature of trust, support real-time expression of comfort or concern, and provide clear and accessible feedback to all occupants. Future research should build on these insights using higher-fidelity environments, synchronized behavioral and trajectory data, and systematic exploration of group dynamics to better understand how collective trust in automated mobility emerges and changes over time.

Acknowledgments

We thank all participants for their time and valuable contributions to this study. We are also grateful to the friends and colleagues who supported the pilot testing phase. The authors extend their appreciation to the members of the lab for their technical assistance and ongoing support throughout the project.

Declaration of competing interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data availability

All data and materials supporting this study are available on the Open Science Framework (OSF): https://osf.io/7d4mw.

The repository includes the baseline trust questionnaire, the manual and automated post-trial questionnaires, the scoring key used to compute composite trust scores, the timeline of critical events for the automated conditions, all processed data used to generate the analyzes and figures, and the full analysis code. All materials are provided exactly as used in the study.

CRediT authorship contribution statement

The authors confirm the following contributions: Conceptualization, Methodology, and Writing – original draft; Data curation, Formal analysis, and Writing – review & editing; Supervision, Methodology, and Resources.

References

[1] National Highway Traffic Safety Administration, National motor vehicle crash causation survey: Report to congress, Tech. Rep. DOT HS 811 059, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC, national Center for Statistics and Analysis (2008).

[2] P. Tafidis, H. Farah, T. Brijs, A. Pirdavani, Safety implications of higher levels of automated vehicles: A scoping review, Transp. Rev. 42 (2) (2022) 245–267. doi:10.1080/01441647.2021.1971794.

[3] O. Tengilimoglu, O. Carsten, Z. Wadud, Implications of automated vehicles for physical road environment: A comprehensive review, Transp. Res. Part E: Logist. Transp. Rev. 169 (2023) 102989. doi:10.1016/j.tre.2022.102989.

[4] S. A. Bagloee, M. Tavana, M. Asadi, T. Oliver, Autonomous vehicles: Challenges, opportunities, and future implications for transportation policies, J. Mod. Transp. 24 (4) (2016) 284–303. doi:10.1007/s40534-016-0117-3.

[5] D. J. Fagnant, K. Kockelman, Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations, Transp. Res. Part A: Policy Pract. 77 (2015) 167–181. doi:10.1016/j.tra.2015.04.003.

[6] F. Biondi, I. Alvarez, K. Jeong, Human–vehicle cooperation in automated driving: A multidisciplinary review and appraisal, Int. J. Hum.-Comput. Interact. 35 (11) (2019) 932–946. doi:10.1080/10447318.2018.1561792.

[7] I. Pettersson, I. C. M. Karlsson, Setting the stage for autonomous cars: A pilot study of future automated driving experiences, IET Intell. Transp. Syst. 9 (7) (2015) 694–701. doi:10.1049/iet-its.2014.0168.

[8] C. Wilson, D. Gyi, A. Morris, R. Bateman, H. Tanaka, Non-driving-related tasks and journey types for future autonomous vehicle owners, Transp. Res.

Part F: Traffic Psychol. Behav. 85 (2022) 150–160. doi:10.1016/j.trf.2022.01.004.

[9] P. Bazilinskyy, M. Kyriakidis, J. de Winter, An international crowdsourcing study into people's statements on fully automated driving, Procedia Manuf. 3 (2015) 2534–2542. doi:10.1016/j.promfg.2015.07.540.

[10] D. Shepardson, Gm self-driving unit cruise admits to submitting false report, will pay $500,000 fine, Reuters, https://tinyurl.com/reuters-cruise-2024 (Nov. 2024).

[11] K. A. Hoff, M. Bashir, Trust in automation: Integrating empirical evidence on factors that influence trust, Hum. Factors 57 (3) (2015) 407–434. doi:10.1177/0018720814547570.

[12] J. K. Choi, Y. G. Ji, Investigating the importance of trust on adopting an autonomous vehicle, Int. J. Hum.-Comput. Interact. 31 (10) (2015) 692–702. doi:10.1080/10447318.2015.1070549.

[13] G. De Cet, F. Orsini, C. Meneguzzer, M. Gastaldi, M. Saljoqi, R. Rossi, Do we trust automated vehicles? a driving simulator study, Transp. Res. Procedia 78 (2024) 174–181. doi:10.1016/j.trpro.2024.02.023.

[14] A. Momen, E. J. de Visser, M. R. Fraune, A. Madison, M. Rueben, K. Cooley, C. C. Tossell, Group trust dynamics during a risky driving experience in a Tesla model x, Front. Psychol. 14 (2023). doi:10.3389/fpsyg.2023.1129369.

[15] J. D. Lee, K. A. See, Trust in automation: Designing for appropriate reliance, Hum. Factors 46 (1) (2004) 50–80. doi:10.1518/hfes.46.1.50_30392.

[16] F. Walker, Y. Forster, S. Hergeth, J. Kraus, W. Payre, P. Wintersberger, M. Martens, Trust in automated vehicles: Constructs, psychological processes, and assessment, Front. Psychol. 14 (2023). doi:10.3389/fpsyg.2023.1279271.

[17] B. M. Muir, Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems, Ergon. 37 (11) (1994) 1905–1922. doi:10.1080/00140139408964957.

[18] R. A. Kaufman, E. Lee, M. S. Bedmutha, D. Kirsh, N. Weibel, Predicting trust in autonomous vehicles: modeling psychosocial traits, risk-benefit attitudes, and driving factors with machine learning, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery (ACM), New York, NY, 2025, pp. 1–24. doi:10.1145/3706598.3713188.

[19] M. Dikmen, C. Burns, Trust in autonomous vehicles: The case of Tesla autopilot and summon, in: IEEE Int. Conf. Syst., Man Cybern. (SMC), 2017, pp. 1093–1098. doi:10.1109/SMC.2017.8122757.

[20] A. Weinstock, T. Oron-Gilad, Y. Parmet, The effect of system aesthetics on trust, cooperation, satisfaction and annoyance in an imperfect automated system, Work 41 (S1) (2012) 258–265. doi:10.3233/WOR-2012-0166-258.

[21] S. Nordhoff, J. Stapel, X. He, A. Gentner, R. Happee, Perceived safety and trust in sae level 2 partially automated cars: results from an online questionnaire, PLoS ONE 16 (12) (2021) e0260953. doi:10.1371/journal.pone.0260953.

[22] Y. Lu, N. Sarter, Eye tracking: A process-oriented method for inferring trust in automation as a function of priming and system reliability, IEEE Trans. Hum.-Mach. Syst. 49 (6) (2019) 560–568. doi:10.1109/THMS.2019.2930980.

[23] J. Ayoub, L. Avetisian, X. J. Yang, F. Zhou, Real-time trust prediction in conditionally automated driving using physiological measures, IEEE Trans. Intell. Transp. Syst. 24 (12) (2023) 14642–14650. doi:10.1109/TITS.2023.3295783.

[24] S. Hergeth, L. Lorenz, R. Vilimek, J. F. Krems, Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving, Human Factors 58 (3) (2016) 509–519. doi:10.1177/0018720815625744.

[25] F. Walker, J. Wang, M. H. Martens, W. B. Verwey, Gaze behaviour and electrodermal activity: Objective measures of drivers' trust in automated vehicles, Transportation Research Part F: Traffic Psychology and Behaviour 64 (2019) 401–412. doi:10.1016/j.trf.2019.05.021.

[26] K. Felbel, A. Dettmann, A. C. Bullinger, Analysis of eye gaze given different automated driving styles in an urban environment, in: Proceedings of the 2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications, 2022. doi:10.1109/CIVEMSA53371.2022.9853647.

[27] C. Strauch, K. Mühl, K. Patro, C. Grabmaier, S. Reithinger, M. Baumann, A. Huckauf, Real autonomous driving from a passenger's perspective: Two experimental investigations using gaze behaviour and trust ratings in field and simulator, Transp. Res. Part F: Traffic Psychol. Behav. 66 (2019) 15–28. doi:10.1016/j.trf.2019.08.013.

[28] Q. Meteier, M. Capallera, E. de Salis, L. Angelini, S. Carrino, M. Widmer, O. Abou Khaled, E. Mugellini, A. Sonderegger, A dataset on the physiological state and behavior of drivers in conditionally automated driving, Data Brief 47 (2023) 109027. doi:10.1016/j.dib.2023.109027.

[29] C. Gold, M. Körber, C. Hohenberger, D. Lechner, K. Bengler, Trust in automation–before and after take-over scenarios in a highly automated vehicle, Procedia Manuf. 3 (2015) 3025–3032. doi:10.1016/j.promfg.2015.07.847.

[30] X. Sun, J. Li, P. Tang, S. Zhou, X. Peng, H. N. Li, Q. Wang, Exploring personalised autonomous vehicles to influence user trust, Cogn. Comput. 12 (6) (2020) 1170–1186. doi:10.1007/s12559-020-09757-x.

[31] R. Thomaschke, C. Haering, Predictivity of system delays shortens human response time, Int. J. Hum.-Comput. Stud. 72 (3) (2014) 358–365. doi:10.1016/j.ijhcs.2013.12.004.

[32] S. Atakishiyev, M. Salameh, R. Goebel, Incorporating explanations into human–machine interfaces for trust and situation awareness in autonomous vehicles, in: Proceedings of the 2024 IEEE Intelligent Vehicles Symposium, 2024, pp. 2948–2955. doi:10.1109/IV55156.2024.10588812.

[33] E. Fu, D. Hyde, S. Sibi, M. Johns, M. Fischer, D. Sirkin, Assessing the effects of failure alerts on transitions of control from autonomous driving systems, in: Proceedings of the 2020 IEEE Intelligent Vehicles Symposium, 2020, pp. 1956–1963. doi:10.1109/IV47402.2020.9304627.

[34] S. Taylor, M. Wang, M. Jeon, Reliable and transparent in-vehicle agents lead to higher behavioral trust in conditionally automated driving systems, Front. Psychol. 14 (2023). doi:10.3389/fpsyg.2023.1121622.

[35] U. Ju, L. L. Chuang, C. Wallraven, Acoustic cues increase situational awareness in accident situations: a vr car-driving study, IEEE Trans. Intell. Transp. Syst. 23 (4) (2022) 3281–3291. doi:10.1109/TITS.2020.3035374.

[36] J. Kraus, D. Scholz, D. Stiegemeier, M. Baumann, The more you know: trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency, Human Factors 62 (5) (2020) 718–736. doi:10.1177/0018720819853686.

[37] S. Baccari, M. Hadded, H. Ghazzai, H. Touati, M. Elhadef, Anomaly detection in connected and autonomous vehicles: a survey, analysis, and research challenges, IEEE Access 12 (2024) 19250–19276. doi:10.1109/ACCESS.2024.3361829.

[38] M. Hossain, M. Abdel-Aty, M. A. Quddus, Y. Muromachi, S. N. Sadeek, Real-time crash prediction models: state-of-the-art, design pathways and

ubiquitous requirements, Accid. Anal. Prev. 124 (2019) 66–84. doi:10.1016/j.aap.2018.12.022.

[39] T. A. G. Venema, F. M. Kroese, J. S. Benjamins, D. T. D. de Ridder, When in doubt, follow the crowd? responsiveness to social proof nudges in the absence of clear preferences, Front. Psychol. 11 (2020) 1385. doi:10.3389/fpsyg.2020.01385.

[40] S. Yoo, J. Kumagai, S. Managi, Trust more, fear less: The role of social support in fully automated vehicle choice, IEEE Trans. Eng. Manag. 71 (2024) 5000–5017. doi:10.1109/TEM.2022.3208567.

[41] S. Dolins, H. Strömberg, Y. Z. Wong, M. Karlsson, Sharing anxiety is in the driver's seat: analyzing user acceptance of dynamic ridepooling and its implications for shared autonomous mobility, Sustainability 13 (14) (2021) 7828. doi:10.3390/su13147828.

[42] F. Faul, E. Erdfelder, A.-G. Lang, A. Buchner, G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behav. Res. Methods 39 (2) (2007) 175–191. doi:10.3758/BF03193146.

[43] P. Bazilinskyy, L. Kooijman, D. Dodou, J. C. F. De Winter, Coupled simulator for research on the interaction between pedestrians and (automated) vehicles, in: Proceedings of the Driving Simulation Conference, Antibes, France, 2020.

[44] J.-Y. Jian, A. M. Bisantz, C. G. Drury, Foundations for an empirically determined scale of trust in automated systems, Int. J. Cogn. Ergon. 4 (1) (2000) 53–71. doi:10.1207/S15327566IJCE0401_04.

[45] L. Lohaus, M. Woide, N. Damm, Z. Demiral, H. Friedrich, A. Petáková, F. Walker, Automated or human: which driver wins the race for the passengers' trust? examining passenger trust in human-driven and automated

vehicles following a dangerous situation, Comput. Hum. Behav. 161 (2024) 108387. doi:10.1016/j.chb.2024.108387.

[46] V. Braun, V. Clarke, Using thematic analysis in psychology, Qual. Res. Psychol. 3 (2) (2006) 77–101. doi:10.1191/1478088706qp063oa.

[47] X. Tan, Y. Zhang, Driver situation awareness for regaining control from conditionally automated vehicles: a systematic review, Human Factors 67 (4) (2025) 367–403. doi:10.1177/00187208241272071.

[48] C. Loersch, H. Aarts, B. K. Payne, V. E. Jefferis, The influence of social groups on goal contagion, J. Exp. Soc. Psychol. 44 (6) (2008) 1555–1558. doi:10.1016/j.jesp.2008.07.009.

[49] J. Domeyer, V. Venkatraman, M. Price, J. D. Lee, Characterizing driver trust in vehicle control algorithm parameters, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 62 (1) (2018) 1821–1825. doi:10.1177/1541931218621413.

[50] T. T. M. Tran, C. Parker, M. Tomitsch, A review of virtual reality studies on autonomous vehicle–pedestrian interaction, IEEE Trans. Hum.-Mach. Syst. 51 (6) (2021) 641–652. doi:10.1109/THMS.2021.3107517.

[51] Y. Guo, X. J. Yang, C. Shi, Tip: a trust inference and propagation model in multi-human multi-robot teams, in: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, 2023, pp. 639–643. doi:10.1145/3568294.3580164.

[52] S. Atakishiyev, M. Salameh, R. Goebel, Incorporating explanations into human–machine interfaces for trust and situation awareness in autonomous vehicles, arXiv preprint arXiv:2404.07383 (2024). doi:10.48550/arXiv.2404.07383.