Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Trust–reliance alignment is context-sensitive in automated driving: a within-person secondary analysis of open experimental data

Xiaoqing Zhao

836693981@qq.com

ZJGS University

Research Article

Keywords: trust in automation, reliance, automated driving, human–autonomy interaction, transparency, within-person analysis

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10843439/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Trust in automation is often summarized as a person-level attitude, although reliance decisions unfold within people as situations change. This study reanalyzed openly available data from 206 licensed drivers who evaluated six scenes involving an automated vehicle, yielding 1,236 repeated observations. Participants viewed either a high- or low-information interface and completed highway and suburban scenarios containing an unexpected event and its resolution. We separated stable between-person trust from scene-specific within-person trust deviations and estimated cluster-robust models of the level of automation participants were willing to confer. Within-person trust change was positively aligned with conferred automation (b = 0.291, 95% CI [0.219, 0.362]), independently of between-person trust, scenario, scene, interface information, order, age, negative attitudes toward automated vehicles, and driver-behavior scores. The same individual conferred substantially less automation in the highway context (b = − 0.612, 95% CI [− 0.714, − 0.509]), and trust–reliance coupling was weaker there (interaction b = − 0.080, 95% CI [− 0.158, − 0.002]). High information did not measurably change that coupling (interaction b = − 0.021, 95% CI [− 0.100, 0.058]). Trust and reliance changes were only modestly correlated across event transitions (Spearman ρ = 0.239–0.324). The findings show that trust level and context-sensitive control allocation are related but non-interchangeable. Interfaces and operating procedures should therefore support timely control reassessment instead of treating higher average trust as the design objective.

1 Introduction

Automated driving turns the allocation of control into recurring cognitive work. A user must interpret what the system is doing, decide how closely to monitor it, and determine whether to continue delegating control as traffic conditions change. This work is not adequately described by a single global rating of trust. A person may report broadly positive trust in an automated vehicle yet reduce the level of automation they are willing to confer when an adjacent vehicle merges dangerously, or may retain a guarded attitude while allowing automation to continue in a routine setting. For safety-critical systems, the consequential question is therefore not simply whether users trust the technology, but when and how their reliance choices adjust to the situation.

Trust has long been treated as an attitude that helps determine willingness to accept vulnerability to automation, whereas reliance is its behavioral manifestation (Lee and See 2004; Hoff and Bashir 2015). Appropriate use requires correspondence between reliance and system capability, not maximal trust (Parasuraman and Riley 1997). The distinction matters operationally because trust attitudes and reliance behavior can diverge. Risk, workload, accountability, interface transparency, and recent system outcomes may all change the cost of acting on a given attitude (Chancey et al. 2017; Hoesterey and Onnasch 2023). Consequently, a between-person correlation between average trust and average reliance cannot reveal whether an individual actually updates control allocation when their trust changes across events.

This distinction is especially important in automated driving. The driver and vehicle form a temporarily coupled cognitive system in which monitoring, interpretation, and control can shift between human and machine. The same automation may be acceptable on an uncomplicated suburban road but invite closer supervision during a high-speed conflict. Information-rich interfaces are intended to support this coordination, but more information can raise perceived transparency without necessarily changing reliance or improving calibration. Studies in human–autonomy teaming show that the effects of transparency depend on situation criticality and that risk can affect trust attitude and trust behavior differently (Hoesterey and Onnasch 2023; Simon et al. 2025). Recent longitudinal evidence likewise emphasizes that aggregate compliance can conceal distinct updating strategies (Simon et al. 2026).

Tanevska et al. (2025a) created a particularly useful open dataset for examining these dynamics. In their online experiment, 206 licensed drivers evaluated three scenes in a highway scenario and three scenes in a suburban scenario. Each sequence included an unexpected event and a resolution. Participants chose actions corresponding to a level of automation (LoA) from 0 to 3 and then reported trust and comfort. The original report compared information conditions, scenarios, and scenes and presented correlations between aggregate trust and LoA. It concluded that trust was relatively stable across scenarios while LoA was strongly context-dependent. Those analyses established the broad effects but did not separate the between-person association (people who generally trust more also delegate more) from the within-person association (a particular person delegates more in scenes where their trust is higher than their own average).

The present secondary analysis addresses that unresolved level-of-analysis problem. It uses person-mean decomposition to estimate stable between-person trust and scene-specific within-person trust simultaneously (Curran and Bauer 2011). This is not a re-test of whether the high-information interface increases average trust—the original paper already addressed that question. Instead, it asks whether trust attitude and reported reliance move together within individuals, whether their coupling changes with driving context or interface information, and whether event-to-event changes in the two measures are aligned. This focus places control allocation, rather than trust maximization, at the center of human–automation design.

The study contributes to Cognition, Technology & Work in three ways. First, it treats the driver’s repeated allocation of control as cognitive coordination with a technological agent. Second, it distinguishes person-level disposition from situation-level updating, avoiding an ecological interpretation of aggregate correlations. Third, it translates that distinction into design and organizational implications for interface information, monitoring support, and the operational handover of control in road systems.

2 Conceptual background and research questions

2.1 Trust attitude, reliance, and level of analysis

Trust in automation is commonly defined in relation to uncertainty, vulnerability, and expectations about system performance (Lee and See 2004). Reliance is the decision to use or defer to automation. These constructs are related but not identical: a favorable attitude can coexist with monitoring or intervention when consequences are serious, and reliance can occur for reasons other than trust, such as workload, convenience, or lack of alternatives. The misuse–disuse framework therefore evaluates automation use against what the technology can actually do rather than against a preferred level of trust (Parasuraman and Riley 1997).

Repeated interactions introduce another distinction. Between-person trust describes stable differences among people; within-person trust describes departures from each person's own typical level in a particular scene. Pooling both sources of variation in one correlation can produce a coefficient that is neither a clean estimate of individual differences nor a clean estimate of dynamic updating. Person-mean decomposition estimates both simultaneously and is recommended whenever a time-varying predictor is used to explain a time-varying outcome (Curran and Bauer 2011). For human–automation research, this distinction maps directly onto two design questions: who tends to rely on automation, and when does the same user revise reliance?

2.2 Situation criticality and transparency

Situational risk changes the consequences of reliance. Evidence from information and decision automation indicates that risk may leave reported trust attitude relatively unchanged while reducing trust behavior or verification omissions (Hoesterey and Onnasch 2023). In human–autonomy teaming, transparency and criticality interact: information that supports cooperation in routine settings can produce different compliance patterns when consequences become severe (Simon et al. 2025). In automated driving, a highway conflict carries speed, proximity, and limited-response-time cues that can motivate control reduction beyond what a global trust score expresses.

Transparency is similarly not a uniform dose. An information-rich interface may improve understanding or perceived trustworthiness, but the effect on behavior depends on whether information is timely, diagnostic, and usable. A static high-information display could raise average trust without strengthening scene-by-scene coupling between trust and reliance. Conversely, it could make reliance more sensitive to trust by helping users interpret why the vehicle acts. Testing the interaction between information condition and within-person trust therefore evaluates whether richer information changes the trust–reliance relationship, not merely the mean of either variable.

2.3 Research questions

- RQ1. After separating stable between-person differences, are scene-specific deviations in trust aligned with the level of automation the same individual is willing to confer?
- RQ2. Does within-person trust–reliance alignment differ between highway and suburban contexts or between high- and low-information interfaces?

- RQ3. Across the onset and resolution of unexpected events, how strongly are changes in trust aligned with changes in conferred automation?

Because the analysis was conceived after the source article and dataset were public, all questions and tests are explicitly exploratory. Estimates and confidence intervals are emphasized; p values are used as graded evidence rather than as a preregistered decision rule.

3 Method

3.1 Study design and data source

We conducted a secondary analysis of the de-identified participant-level dataset deposited by Tanevska, Winkle, and Castellano (2025b) on Zenodo (https://doi.org/10.5281/zenodo.15648216). The dataset is licensed under Creative Commons Attribution 4.0 and accompanies the published experimental report (Tanevska et al. 2025a). No new participants were recruited and no simulated or imputed participant records were used.

The source study used a 2 × 2 between-participant design: high versus low interface information and highway-first versus suburbs-first scenario order. Each participant evaluated both scenarios, each comprising three scenes. In the highway scenario, another vehicle merged at unsafe distance and speed in Scene 2 and the automated vehicle then braked. In the suburban scenario, the vehicle unexpectedly slowed in Scene 2 and a previously obscured school became visible in the resolution. The high-information interface displayed processed environmental information and the vehicle's recommended action; the low-information interface displayed navigation and speed information. Full stimuli and procedures are described in the source article (Tanevska et al. 2025a).

3.2 Participants and source-study quality controls

The source study recruited English-speaking adults through Prolific who held a valid driver's license. Of 249 initial responses, 43 were excluded by the source investigators for failed attention/validity checks or implausible completion time, leaving 206 participants. The released dataset contains these 206 complete cases. Age ranged from 20 to 73 years (M = 36.42, SD = 13.18); recorded sex was equally divided between female and male participants. The high-information condition included 102 participants and the low-information condition 104. There were 105 highway-first and 101 suburbs-first participants.

Table 1
Participant and design characteristics

Characteristic	Value
Participants	206
Repeated scene observations	1,236
Age, years	M = 36.42, SD = 13.18, range 20–73
Sex	103 female; 103 male
Interface information	102 high; 104 low
Scenario order	105 highway first; 101 suburbs first
Active driving frequency	136 daily; 59 a few days/week; 11 less often or no longer driving

Note. Counts and descriptive statistics were computed from the publicly released complete-case file.

3.3 Measures

Conferred level of automation (reported reliance choice). After each scene, participants selected one or more actions in response to “What do you do?” Actions ranged from taking over driving control (SAE Level 0) to allowing eyes-off automation (Level 3). When multiple actions were selected, the source investigators averaged their coded levels. The scene score therefore ranged continuously from 0 to 3, with higher values indicating willingness to confer more automation. This is a stated action choice in a vignette, not observed on-road behavior.

Trust attitude. After the action choice, participants selected statements describing whether the vehicle was reliable, safe, and worthy of reliance or, conversely, unsafe or untrustworthy. Positive and negative selections were summed by the source investigators to a score from − 2 to 4. Higher scores indicate more positive scene-specific trust. Because trust was recorded after the action choice, analyses quantify contemporaneous alignment and do not treat trust as a temporally prior causal mediator.

Comfort and individual-difference covariates. Scene comfort was coded − 1 (anxious/tense), 0 (neutral/unsure), or 1 (relaxed/calm). Participant-level covariates were age, an adapted 10-item Driver Behaviour Questionnaire (DBQ; 0–5 average), and a 10-item automated-vehicle adaptation of the Negative Attitudes toward Robots Scale (AV-NARS; 1–5 average). The source article reported internal consistency of α = .772 for DBQ and α = .854 for AV-NARS.

3.4 Data preparation and analysis

The 206 × 153 wide-format file was reshaped to one row per participant, scenario, and scene (1,236 rows). Participant identifiers were replaced in the derived file with sequential analytic labels (P001–P206). No observations were dropped and no missing-value imputation was needed for variables used in the analysis.

For participant i in scene j, trust was decomposed into the participant's six-scene mean (between-person trust) and the deviation of the scene score from that mean (within-person trust). Conferred LoA was modeled by ordinary least squares with standard errors clustered by participant (Cameron and Miller 2015). Model 1 included between- and within-person trust, highway context, Scene 2 and Scene 3 indicators (Scene 1 reference), high-information interface, first scenario exposure, interactions of within-person trust with highway context and high information, and standardized age, AV-NARS, and DBQ scores. Model 2 added between- and within-person comfort as a sensitivity test of whether trust alignment remained after concurrent affective state was considered.

Unstandardized coefficients are reported because both focal scales have meaningful ranges. A coefficient for within-person trust is the expected change in LoA for a one-point scene-specific increase in trust relative to that participant's own mean, holding the listed covariates constant. Conditional slopes were calculated as linear combinations of model coefficients. For event transitions (Scene 1→2 and Scene 2→3 in each scenario), Spearman correlations related within-person trust changes to LoA changes; 95% percentile intervals used 5,000 participant bootstrap resamples with seed 20260826. Four transition p values were adjusted by Holm's procedure. A standardized version of Model 1 was also estimated. The complete script, derived data, numerical output, and audit workbook are supplied as Online Resources 1–3.

All analyses used Python 3 with pandas, NumPy, SciPy, and Matplotlib. Cluster-robust covariance followed the CR1 finite-sample correction, with 205 cluster degrees of freedom. The deterministic analysis script validates the expected 206 × 153 source shape before computing results.

3.5 Ethics, data governance, and generative-AI assistance

The original investigators reported that the Swedish Ethical Review Authority (Etikprövningsmyndigheten; application 2023-01688-01) evaluated the project as not requiring ethical approval under Section 4 of the Swedish Ethical Review Act; participants provided written informed consent and no minors or vulnerable participants were involved (Tanevska et al. 2025a). The present study analyzed only de-identified data already released publicly under CC BY 4.0 and involved no new participant contact. The derived file omits the source's free-text fields and replaces source identifiers with sequential labels.

OpenAI ChatGPT (Codex environment; accessed 26 August 2026) assisted with retrieval planning, analysis-code drafting, manuscript organization, and English-language drafting. Numerical results in this article were produced by the supplied deterministic script from the cited dataset. Consistent with Springer Nature policy, the AI system is not an author; the submitting human author must verify the analysis, citations, and final wording and accepts responsibility for the submitted work.

4 Results

4.1 Scene-level trajectories

Both trust and conferred LoA were lower in the highway scenario and changed markedly around unexpected events (Fig. 1; Table 2). In the highway sequence, mean trust declined from 1.451 in Scene 1 to 0.796 in Scene 2 and 0.587 in Scene 3; mean LoA declined from 1.900 to 1.071 and then 0.820. In the suburban sequence, mean trust changed from 1.748 to 1.558 and recovered to 1.830; mean LoA declined from 2.429 to 1.837 and recovered slightly to 1.921. The trajectories therefore show parallel movement but also a notable asymmetry: the suburban trust recovery was larger than the LoA recovery, whereas both measures continued downward in the highway resolution.

Table 2
Scene-level means and standard deviations

Scenario	Scene	Trust M (SD)	LoA M (SD)	Comfort M (SD)
Highway	1	1.451 (1.924)	1.900 (0.809)	-0.044 (0.773)
Highway	2	0.796 (1.891)	1.071 (0.885)	-0.383 (0.741)
Highway	3	0.587 (1.853)	0.820 (0.805)	-0.718 (0.599)
Suburbs	1	1.748 (1.698)	2.429 (0.603)	0.437 (0.728)
Suburbs	2	1.558 (1.703)	1.837 (0.891)	0.316 (0.734)
Suburbs	3	1.830 (1.643)	1.921 (0.988)	0.510 (0.675)
Note. N = 206 at every scenario–scene combination. Scene 2 contained the unexpected event.				

4.2 Within-person trust–reliance alignment

Model 1 explained 45.6% of variation in conferred LoA (Table 3). Between-person trust was positively associated with LoA (b = 0.205, 95% CI [0.152, 0.259], p < .001): participants with a one-point higher six-scene mean trust conferred about 0.21 more LoA units on average. Independently, within-person trust was positively aligned with LoA in the reference condition (suburban, low-information; b = 0.291, 95% CI [0.219, 0.362], p < .001). Thus, when the same participant's trust was one point above their own mean, their reported action corresponded to approximately 0.29 more LoA units, conditional on design factors and covariates.

The highway context had a large negative association with LoA even after trust was decomposed and controlled (b = − 0.612, 95% CI [− 0.714, − 0.509], p < .001). It also attenuated the within-person trust slope by 0.080 LoA units per trust point (95% CI [− 0.158, − 0.002], p = .044). Conditional slopes remained positive in both contexts: 0.291 in suburbs and 0.210 on the highway under low information. The high-information interface did not measurably alter the within-person slope (interaction b = − 0.021, 95% CI [− 0.100, 0.058], p = .599). Across the four scenario × information combinations, conditional slopes ranged from 0.189 to 0.291 and were all positive (all p < .001; Online Resource 1). The evidence therefore supports dynamic alignment but also shows that situation affects control allocation beyond concurrent trust attitude.

Table 3
Cluster-robust models of conferred level of automation

Term	M1 b	95% CI	p	M2 b	95% CI	p
Trust: between person	0.205	[0.152, 0.259]	< .001	0.204	[0.148, 0.259]	< .001
Trust: within person	0.291	[0.219, 0.362]	< .001	0.262	[0.188, 0.336]	< .001
Highway context	-0.612	[-0.714, -0.509]	< .001	-0.505	[-0.623, -0.388]	< .001
Scene 2	-0.616	[-0.699, -0.532]	< .001	-0.591	[-0.678, -0.505]	< .001
Scene 3	-0.719	[-0.811, -0.627]	< .001	-0.682	[-0.775, -0.590]	< .001
High-information interface	-0.003	[-0.123, 0.116]	= .955	-0.006	[-0.127, 0.115]	= .919
First scenario exposure	-0.093	[-0.185, -0.000]	= .050	-0.057	[-0.149, 0.034]	= .218
Within trust × highway	-0.080	[-0.158, -0.002]	= .044	-0.085	[-0.162, -0.007]	= .034
Within trust × high information	-0.021	[-0.100, 0.058]	= .599	-0.015	[-0.092, 0.063]	= .709
Age (standardized)	0.039	[-0.021, 0.100]	= .203	0.040	[-0.021, 0.102]	= .197
AV-NARS (standardized)	-0.024	[-0.088, 0.039]	= .447	-0.017	[-0.088, 0.054]	= .640
DBQ (standardized)	-0.036	[-0.101, 0.030]	= .283	-0.037	[-0.101, 0.027]	= .254

Note. N = 1,236 observations nested within 206 participants. M1 R^2 = .456; M2 R^2 = .463. M2 additionally adjusts for between- and within-person comfort. Reference categories are suburban context, Scene 1, low information, and second scenario exposure. Full coefficients appear in Online Resource 1.

4.3 Event-transition alignment

Changes in trust and LoA were positively but only modestly aligned at all four transitions (Table 4). During the highway unexpected event, trust declined by 0.655 points and LoA by 0.829 units on average; their change scores correlated at ρ = .308. Both declined again during the highway resolution (ρ = .324). In suburbs, the unexpected event produced a relatively small mean trust decline (− 0.189) but a larger LoA decline (− 0.591), with ρ = .239. Trust then recovered by 0.272 points while LoA recovered by only 0.084, with ρ = .309. All four correlations remained statistically distinguishable from zero after Holm

adjustment. Their moderate magnitude indicates that attitude and action-choice updating shared variance but were far from interchangeable.

Table 4
Alignment of trust and LoA changes across scene transitions

Scenario	Transition	Δ Trust M	Δ LoA M	Spearman ρ	95% boot CI	Holm p
Highway	Event (1→2)	-0.655	-0.829	0.308	[0.182, 0.425]	< .001
Highway	Resolution (2→3)	-0.209	-0.251	0.324	[0.182, 0.451]	< .001
Suburbs	Event (1→2)	-0.189	-0.591	0.239	[0.091, 0.374]	< .001
Suburbs	Resolution (2→3)	0.272	0.084	0.309	[0.168, 0.443]	< .001

Note. Positive change indicates an increase from the earlier to the later scene. Confidence intervals are participant-bootstrap percentile intervals (5,000 resamples).

4.4 Sensitivity analyses

Adding between- and within-person comfort increased explained variation to 46.3% but did not materially change the focal estimates. Within-person trust remained positively aligned with LoA (b = 0.262, 95% CI [0.188, 0.336], p < .001), and the highway attenuation remained evident (interaction b = − 0.085, 95% CI [− 0.162, − 0.007], p = .034). The information interaction remained near zero (b = − 0.015, 95% CI [− 0.092, 0.063], p = .709). In the standardized Model 1, the between-person trust coefficient was 0.293 and the within-person coefficient 0.340 standard deviations. Age, AV-NARS, and DBQ did not have statistically distinguishable conditional associations in Model 1. Complete output is provided in Online Resource 1.

5 Discussion

5.1 Main interpretation

The analysis supports a simple but consequential proposition: trust and reliance move together within people, yet situation-specific control allocation cannot be reduced to trust attitude. Participants who generally trusted the automated vehicle more also conferred more automation. More importantly, a participant conferred more automation in scenes where their own trust was higher than usual. This within-person relationship remained after scenario structure, interface condition, order, age, negative attitudes toward automated vehicles, driver-behavior scores, and concurrent comfort were considered.

At the same time, highway context reduced conferred automation by more than half a LoA unit after these factors were controlled and weakened the within-person trust slope. In other words, participants did not simply translate a given trust score into the same control choice everywhere. In the higher-speed

highway context, they tended to retain more control regardless of their momentary trust. This is consistent with the broader argument that vulnerability changes how an attitude is converted into behavior (Chancey et al. 2017; Hoesterey and Onnasch 2023). The present finding adds a within-person automated-driving test: contextual criticality shifted both the level of reliance and its coupling with self-reported trust.

The transition analysis reinforces that conclusion. Trust and LoA changes were aligned, but correlations of .24 to .32 leave most variation unshared. The asymmetry was particularly clear in suburbs: the unexpected slowing event triggered a much larger control reduction than trust reduction, and the explanation/resolution restored trust more strongly than LoA. That pattern is compatible with cautious monitoring after ambiguity has been explained. It also shows why a recovered trust rating should not be read as full restoration of willingness to delegate control.

5.2 Implications for human–automation design and work

The primary design implication is to optimize control reassessment rather than average trust. A dashboard that increases mean trust may still fail to help users decide when to monitor, intervene, or resume delegation. Interfaces should expose information that is diagnostic for the current control decision: detected hazard, confidence or uncertainty, intended maneuver, operating boundary, and the action expected from the driver. The information should be timed around state transitions, not merely made permanently abundant.

The null interaction with high information is informative but should be interpreted cautiously. The estimated difference in within-person coupling was small and imprecise enough to include modest effects in either direction. Still, the data do not support the claim that displaying a fuller set of processed information and a recommended action automatically makes trust and reliance more tightly coordinated. Information quantity is not equivalent to decision support. Designers should test whether information improves recognition of when control should shift, not just whether users report that the system seems more transparent.

For organizations that deploy automated driving, the results imply a monitoring responsibility. Evaluation dashboards should report trust attitude and reliance behavior separately and should stratify both by scenario criticality. Aggregate trust or average compliance can conceal risky over-reliance in some situations and unnecessary disuse in others. Training and interface evaluation should include event onset and resolution because recovery is not a single process: users may cognitively understand an explanation before they are willing to restore control to automation. Recent CTW work similarly shows that average compliance can hide distinct longitudinal strategies (Simon et al. 2026).

The cognitive-work perspective also reframes the driver's role. Even when the vehicle performs the dynamic driving task, the human may still carry responsibility for interpreting boundaries, supervising uncertainty, and taking over. That work must be supported by organizational procedures as well as interface elements: unambiguous handover protocols, consistent alert semantics, post-event

explanations, and training that links cues to actions. Poorly designed transparency can transfer interpretive burden to the driver without improving control decisions.

5.3 Strengths and limitations

A major strength is the use of participant-level open data with six repeated observations per person. The within–between decomposition directly matches the theoretical distinction between stable propensity and scene-specific updating, and clustered inference respects the dependence of repeated observations. The derived data and complete script make every reported coefficient, interval, and figure reproducible.

Several limitations constrain inference. First, the source study used static online vignettes. Conferred LoA is a self-reported action choice, not observed takeover behavior, and the scenarios cannot reproduce vehicle motion, time pressure, workload, or physical vulnerability. Second, trust was measured after the action choice, so the models establish alignment rather than a causal pathway from trust to reliance. Third, the trust score was assembled from selected statements and is not a standardized continuous scale. Fourth, each scenario had only three scenes and combined multiple contextual differences; “highway” is not a randomized pure-risk manipulation. Fifth, the sample was recruited through Prolific, required English fluency and a license, and may not generalize to professional drivers, non-drivers, or specific national populations. Sixth, because this secondary analysis was not preregistered, p values are exploratory and the highway interaction should be replicated.

Future studies should record trust attitude before action as well as after feedback, obtain objective reliance and takeover measures in simulators or controlled vehicles, manipulate reliability and consequence separately, and sample enough events to estimate person-specific dynamics. A strong design would compare static information quantity with event-triggered, action-oriented transparency and would define an external criterion for appropriate reliance based on verified system capability. Such a criterion is necessary before the present alignment measure can be called calibration in the normative sense.

6 Conclusion

In open experimental data from 206 drivers, trust attitude and reported reliance were aligned both between people and within the same person across automated-driving scenes. Yet the relationship was context-sensitive: highway scenes reduced delegation beyond concurrent trust and weakened the coupling between trust and control allocation, while a richer information interface did not measurably strengthen that coupling. Event transitions further showed only modest correspondence between trust change and reliance change. Human–automation systems should therefore evaluate and support the dynamics of control allocation rather than treating higher average trust as success. The operational question is not only whether users trust an automated vehicle, but whether they can translate changing evidence and risk into timely, proportionate control decisions.

Declarations

Funding. Author confirmation is required on the separate title-page form before submission; no funding information was available in the supplied materials.

Competing interests. Author confirmation is required on the separate title-page form before submission; no competing-interest information was available in the supplied materials.

Ethics approval. No new human-participant research was conducted. The source investigators reported that Etikprövningsmyndigheten (application 2023-01688-01) evaluated the original project as not requiring ethical approval under Section 4 of the Swedish Ethical Review Act.

Consent to participate. The source investigators reported written informed consent from all participants.

Consent for publication. Not applicable; this secondary analysis reports no identifiable individual information.

Data availability. The de-identified source data are publicly available from Zenodo at https://doi.org/10.5281/zenodo.15648216 under CC BY 4.0. The analysis-ready derived file and complete analysis script are supplied as Online Resources 2 and 3.

Code availability. The exact Python script used to reshape the source data, estimate all models, create tables and figures, and write machine-readable results is supplied as Online Resource 2.

Author contributions. Xiaoqing Zhao: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing—original draft, Writing—review and editing. The submitting author must verify and, if necessary, amend this statement before submission.

References

1. Cameron AC, Miller DL (2015) A practitioner's guide to cluster-robust inference. J Hum Resour 50(2):317–372. https://doi.org/10.3368/jhr.50.2.317
2. Chancey ET, Bliss JP, Yamani Y, Handley HA (2017) Trust and the compliance–reliance paradigm: the effects of risk, error bias, and reliability on trust and dependence. Hum Factors 59(3):333–345. https://doi.org/10.1177/0018720816682648
3. Curran PJ, Bauer DJ (2011) The disaggregation of within-person and between-person effects in longitudinal models of change. Annu Rev Psychol 62:583–619. https://doi.org/10.1146/annurev.psych.093008.100356
4. de Visser EJ, Peeters MMM, Jung MF, Kohn S, Shaw TH, Pak R, Neerincx MA (2020) Towards a theory of longitudinal trust calibration in human–robot teams. Int J Soc Robot 12:459–478. https://doi.org/10.1007/s12369-019-00596-x
5. Hoff KA, Bashir M (2015) Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 57(3):407–434. https://doi.org/10.1177/0018720814547570

6. Hoesterey S, Onnasch L (2023) The effect of risk on trust attitude and trust behavior in interaction with information and decision automation. Cogn Technol Work 25:15–29. https://doi.org/10.1007/s10111-022-00718-y
7. Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. Hum Factors 46(1):50–80. https://doi.org/10.1518/hfes.46.1.50_30392
8. Parasuraman R, Riley V (1997) Humans and automation: use, misuse, disuse, abuse. Hum Factors 39(2):230–253. https://doi.org/10.1518/001872097778543886
9. Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern A 30(3):286–297. https://doi.org/10.1109/3468.844354
10. Simon L, Rauffet P, Guerin C (2025) Effects of agent transparency and situation criticality upon human-autonomy trust and risk perception in decision-making. Cogn Technol Work 27:21–38. https://doi.org/10.1007/s10111-024-00782-6
11. Simon L, Rauffet P, Guérin C (2026) Dynamic transparency and feedback effect on operator's trust calibration and compliance in a simulated longitudinal cooperation. Cogn Technol Work. https://doi.org/10.1007/s10111-026-00890-5
12. Tanevska A, Winkle K, Castellano G (2025a) Information vs situation: balancing transparency and autonomy for trustworthy autonomous vehicles. Front Robot AI 12:1657857. https://doi.org/10.3389/frobt.2025.1657857
13. Tanevska A, Winkle K, Castellano G (2025b) User data for paper 'Blending Participatory Design and Artificial Awareness for Trustworthy Autonomous Vehicles' [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.15648216
14. Wildman JL, Nguyen D, Thayer AL, Robbins-Roth VT, Carroll M, Carmody K, Ficke C, Akib M, Addis A (2024) Trust in human-agent teams: a multilevel perspective and future research agenda. Organ Psychol Rev. https://doi.org/10.1177/20413866241253278

Figures

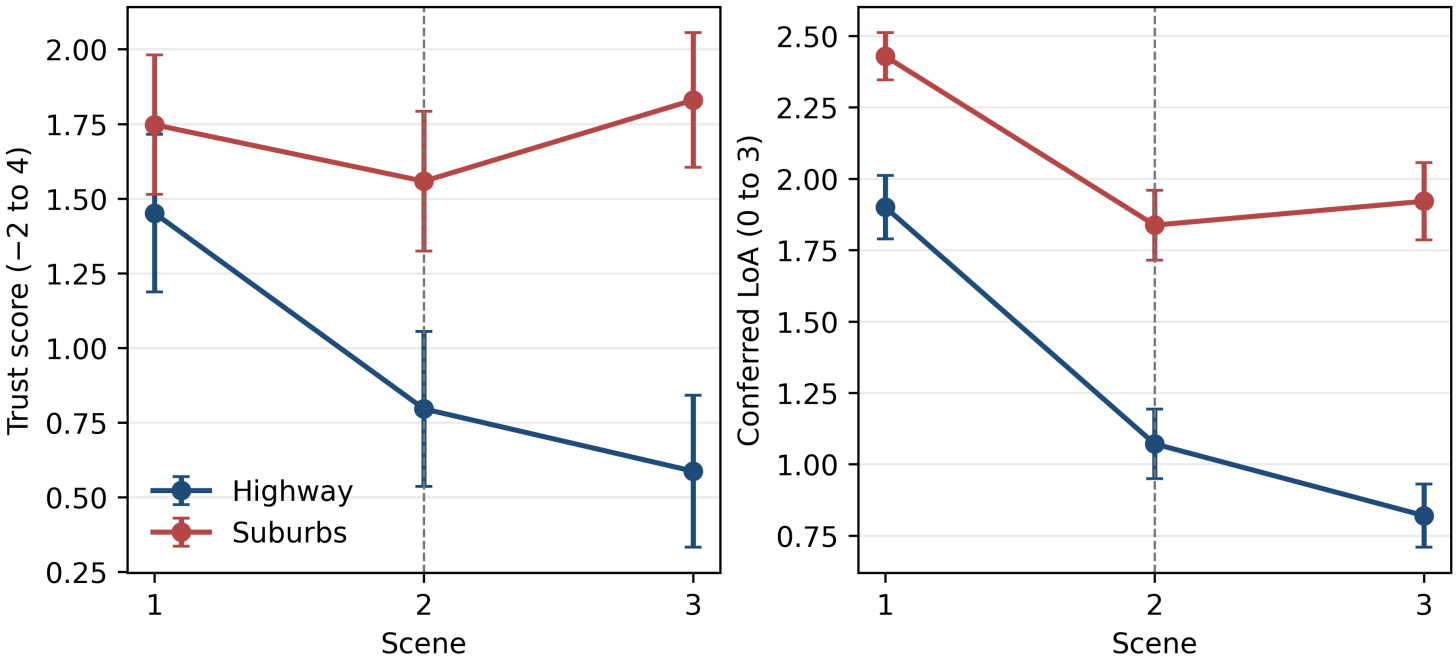


Figure 1

Mean trust and conferred level of automation across scenes. Error bars are 95% confidence intervals for scene means. The dashed line marks the unexpected event in Scene 2. Trust ranged from −2 to 4; conferred LoA ranged from 0 to 3.

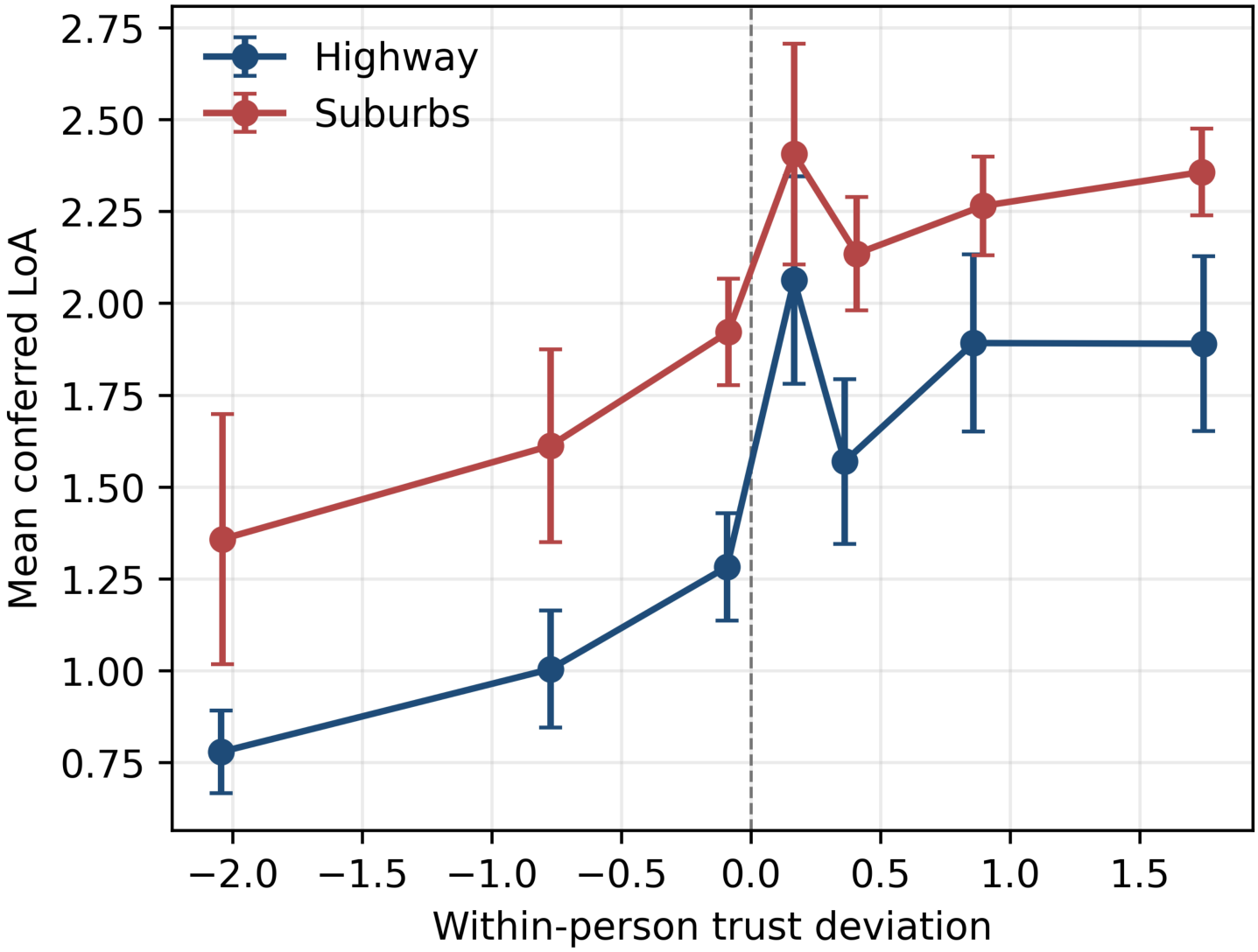


Figure 2

Binned observed relationship between within-person trust deviations and conferred LoA by scenario. Points are means within quantile bins; error bars are approximate 95% confidence intervals. Regression estimates and inference use all 1,236 observations with participant-clustered standard errors, not the bins.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 03OnlineResource2ReproducibleAnalysis.py
- 05OnlineResource3bAnalysisWorkbook.xlsx
- 02OnlineResource1SupplementaryMethods.pdf
- 04OnlineResource3aAnalysisReadyData.csv

- OnlineResourceslegend.docx