Promises and Trust Repair in UGVs

Connor Esterwood[1], Arsha Ali[2], Zariq George[2], Samantha Dubrow[3], Jonathon Smereka[4], Kayla Riegner[4], Dawn Tilbury[2], Lionel P. Robert Jr [1,2]

[1]School of Information, University of Michigan; [2]Robotics Department, University of Michigan; [3]The MITRE Corporation, [4]U.S. Army DEVCOM GVSC

Unmanned ground vehicles (UGVs) are autonomous robots capable of performing tasks through self-navigation and decision-making. They have the potential to replace humans in dangerous driving scenarios. However, UGVs must be viewed as trustworthy to be accepted, and like any automation, they can make mistakes that decrease human trust in them. Trust repair strategies can mitigate the consequences of trust violations, but they are not always effective. To better understand their effectiveness on UGVs, we designed a between-subjects study examining promises on a UGV's trustworthiness. Preliminary results showed that promises had a marginal impact on overall trustworthiness but were influential in repairing benevolence but not ability or integrity. These findings have implications for the design of UGV's and trust repair theory.

INTRODUCTION

Unmanned ground vehicles (UGVs) are autonomous ground robots of various sizes that are capable of performing a shared task through self-navigation and decision making. The use of these vehicles – sometimes also referred to as self-driving or autonomous vehicles (AVs) – have the potential to not only replace human users in dangerous or hazardous driving-related scenarios but also work alongside them in human-robot teams (Ni, Hu, & Xiang, 2021; Rossiter, 2020), in turn, reducing staffing costs and augmenting existing staff capabilities in various industries (Ni et al., 2021). As a result, UGVs have been increasingly adopted in a variety of domains including aiding first responders and various military units (Ni et al., 2021).

UGVs must be viewed as trustworthy to be accepted; yet, like any automation, they can and do fail. Previous research has shown that humans do not accept or ultimately rely on UGVs they do not trust (Ghazizadeh, Lee, & Boyle, 2012; Hancock et al., 2011; Mühl et al., 2020; Parasuraman & Riley, 1997). However, like human teammates, UGVs are not perfect and mistakes happen. These mistakes decrease human's trust in them and, if left unaddressed, can have disastrous effects on human-robot team performance (Baker, Phillips, Ullman, & Keebler, 2018; Sebo, Krishnamurthi, & Scassellati, 2019). This is especially problematic in scenarios where UGVs are meant to act as not only useful tools but also as active collaborators and teammates (Azevedo-Sa et al., 2021; Rossi, Dautenhahn, Koay, & Walters, 2020; Surendran, Mokhtari, & Wagner, 2021; Xu & Howard, 2020; You & Robert Jr, 2018).

However, there are trust repair strategies to mitigate the negative consequences of trust violations, but they generally have shown mixed results (Esterwood & Robert, 2022). Trust repair strategies are actions taken by a trustor to help restore trust after a trust violation (Kramer & Lewicki, 2010). To date, human-robot trust repairs have all been derived from the literature on human-human interactions (Esterwood & Robert, 2022; Lewicki & Brinsfield, 2017). Unfortunately, the efficacy of repair strategies has generally shown mixed results (Esterwood & Robert, 2022). This is especially the case in the context of UGVs (Feng & Tan, 2022; Kohn, Momen, Wiese, Lee, & Shaw, 2019; Kohn, Quinn, Pak, De Visser, & Shaw, 2018; Xu & Howard, 2022). This limits our understanding of what and when repairs may be effective in the context of UGVs.

To better understand the effectiveness of trust repair strategies on UGVs, we designed a between-subjects study with the goal of examining the impact of promises on a UGV's trustworthiness. Trustworthiness is defined as the degree to which a trustee sees a trustor as worthy of their trust and is composed of three sub-components: ability, integrity, and benevolence (Mayer, Davis, & Schoorman, 1995; L. P. Robert, Denis, & Hung, 2009). Unlike previous studies on UGVs, this study focuses on trustworthiness rather than trust. This is because trustworthiness is a precursor to trust and would allow us to explore not only if a particular repair strategy is effective or not but also why. This study also focused on promises as a repair strategy because, although studied less frequently, promises have been shown to be effective in restoring trust (Esterwood & Robert, 2022). Results based on a preliminary analysis found that promises had a marginally significant impact on overall trustworthiness. A closer examination of trustworthiness's sub-components revealed that promises were influential in repairing benevolence. Similar results, however, for the trustworthiness sub-components of ability and integrity were not present.

The paper's preliminary findings contribute to the literature in several ways. First, the findings provide unique insight into the potential efficacy of promises as a repair strategy. Although our findings are preliminary, they highlight that promises may be an effective approach to restoring trust in UGVs through repairing perceptions of trustworthiness. This finding is novel and especially important as other repair strategies have been shown to be primarily ineffective (Esterwood & Robert, 2022; Feng & Tan, 2022; Xu & Howard, 2022). Second, this paper offers contributions to theory by identifying the role of benevolence. This is an important finding as it not only reinforces similar findings in other domains (Esterwood & Robert, 2021, 2023) but also

Distribution A: Approved for public release; distribution unlimited OPSEC #7633.

Approved for Public Release; Distribution Unlimited. Public Release Case Number 23-1782. The authors' affiliation with The MITRE Corporation is ~~pro-vided~~ for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or view-points expressed by the author.

Esterwood, C., Ali, A., George, Z., Dubrow, S., Smereka, J., Riegner, K., Tilbury, D. and Robert, L.P. (2023). **Promises and Trust Repair in UGVs**, *Proceedings of the HFES 67th Annual Meeting of the Human Factors and Ergonomics Society* (**HFES 2023**), October 23-27, 2023, Washington, DC.

highlights a potentially important pathway through which other trust repairs may influence trust.

RELATED WORK

Trust & Trustworthiness

Trust is defined as the willingness to be vulnerable to the actions of another based on the *expectation* that the other will come through (Mayer et al., 1995) where as reliability is the consistency of this other to perform as expected (Johns, 1996). Trust is preceded by trustworthiness which influences the expectations that one has of a trustee (Mayer et al., 1995; L. Robert & You, 2013; Ye et al., 2019). In this way, trustworthiness directly influences when one is willing to bestow trust. Generally, trustworthiness can be divided into three different sub-components: ability, benevolence, and integrity (Lee & Moray, 1992; Mayer et al., 1995; L. P. Robert et al., 2009). Ability –sometimes referred to as performance– is the perceptions of a trustee's skillfulness and competence (Mayer et al., 1995). Integrity –sometimes referred to as process– encompasses perceptions of a trustee's moral consistency and honesty (Kim, Kim, Lyons, & Nam, 2020). Benevolence –sometimes referred to as purpose– is the extent to which the trustee is perceived as caring for the trustor's best interests and acting accordingly (Mayer et al., 1995). In summary, the three sub-components of trustworthiness – ability, integrity, and benevolence – each make up trustworthiness, which in turn, influences one's willingness to bestow trust not only to humans but also to robots (Esterwood & Robert, 2023; Kim et al., 2020; Lyons, aldin Hamdan, & Vo, 2023; Sebo et al., 2019).

Trust & Trust Violations in UGVs

Trust is vital for the adoption of UGVs (Azevedo-Sa et al., 2020; Du et al., 2019; Liu, Guo, Ren, Wang, & Xu, 2019; Liu, Yang, & Xu, 2019; Yokoi & Nakayachi, 2021; Zhang et al., 2019). The more that humans trust UGVs, the more likely it is that they will accept them (Yokoi & Nakayachi, 2021). UGVs, however, are not perfect and failures will eventually occur. These failures –i.e. trust violations– can have dire consequences. In particular, recent examples across the human-factors literature have highlighted how trust violations can negatively impact not only trust (Azevedo-Sa et al., 2020; Hancock et al., 2011; Lyons et al., 2023) but also acceptance (Seet et al., 2020) and task performance – in terms of secondary task – (Azevedo-Sa et al., 2020; Chen & Terrence, 2009) alongside takeover performance (Körber, Baseler, & Bengler, 2018; Payre, Cestac, & Delhomme, 2016). In light of these findings, it becomes increasingly important to consider how trust repairs might be capable of mitigating these negative consequences.

Trust Repairs in UGVs

Currently, there are only a handful of studies that have investigated the efficacy of trust repair strategies in the context of UGVs (Feng & Tan, 2022; Kohn et al., 2019, 2018; Xu & Howard, 2022). Generally, these studies have produced mixed and leaning towards non-significant results. In particular, three studies found that apologies and denials were ineffective in restoring trust (Feng & Tan, 2022; Kohn et al., 2018; Xu & Howard, 2022), two studies found that explanations were ineffective (Feng & Tan, 2022; Kohn et al., 2018), and one study found that combining explanations with promises was ineffective (Feng & Tan, 2022). Conversely, however, one additional study found that both apologies and denials were effective in restoring trust in UGVs (Kohn et al., 2019).

Notably, none of these studies examined the effectiveness of promises in isolation, nor did they examine trustworthiness. There are two gaps across the UGV trust repair literature. First, the current literature fails to provide insight into whether promises can restore trust as an independent repair strategy or if promises are only effective when combined with other repair strategies. Second, by examining trust as a singular overall construct instead of trustworthiness, any differential impact of repairs on ability, integrity, or benevolence remains unexamined. Scholars do not understand how trust repairs function and through what trustworthiness elements these repairs influence trust. In response to these gaps, this paper reports the preliminary results of an ongoing research study comparing trustworthiness in UGVs after two errors and either a promise or a no repair condition. This study seeks to build upon previous work conducted on trust repair in HRI (Esterwood & Robert, 2021, 2023) by examining the specific trust repair strategy of promises in isolation while also focusing on the context area of UGVs as opposed to robots in general.

METHODOLOGY

Apparatus

This study used an immersive virtual environment (IVE) deployed in an in-person laboratory. This IVE required subjects to collaborate with two UGVs while performing a secondary task. To accomplish this, subjects were positioned within a driving simulator and shown a unique graphic user interface (GUI). This GUI was divided into three columns where the far left and right columns showed the status of both the first and second UGV as well as a "live" feed from the UGVs' camera. The middle column of the interface contained the secondary task where subjects attempted to move a box around a specified play area to score points by colliding the box with falling shapes.

During the study, subjects were tasked with monitoring the statuses of the UGVs via the "live" feeds and notification elements while trying to gain as many points as possible in the secondary task. Subjects were compensated a base amount of $25 USD for participation. Bonuses of up to an additional $15 USD were available for subjects who performed well on the secondary task and minimized the amount of time that a UGV spent waiting for a takeover. This encouraged participation and engagement in the study as well as ensured that any errors the UGVs made would be consequential.

Over the course of the study, participants were presented with six takeover requests in the form of visual and auditory notifications. The timing and frequency of these requests were not disclosed, but participants were informed that the requests would be made if the UGV encountered an obstacle on its route to its destination. When participants received a takeover request, they assumed control of the UGV facing the obstacle, and their screen changed to a view from behind the steering wheel. They manually navigated the UGV around the obstacle and returned it to autonomous mode once a dashboard notification indicated that autonomy was available. After completing the takeover, participants resumed their secondary task, which involved monitoring and scoring points to earn bonus payments.

During this study the UGVs make mistakes at 2 points: before the 5th and 7th obstacles. These mistakes took the form of a UGV stopping unexpectedly where no obstacle was present and failing to ask for a takeover. Given that subjects were told that the UGVs were supposed to be capable of autonomous driving and would ask for help only when they needed it, this behavior was expected to violate the human's trust in the UVGs. These trust violations occurred once for each of the two UGVs and allowed us to examine the effects of different trust repair strategies across the course of the study.

Experimental Design

This study employed a between-subjects study comprised of one experimental condition and one control condition with 10 subjects per condition. The experimental condition was comprised of the UGVs offering promises (*promise* condition) after violating the human's trust. Specifically, the UGVs stated "I promise I'll ask for help when I need it next time". Promises occurred only once for each UGV preventing them from violating their promise. The control condition was comprised of the robot remaining silent after violating the human's trust (*no repair* condition). By conducting the experiment in this manner we were able to compare trustworthiness between these two conditions and examine what impact *promises* had on trustworthiness and its sub-components.

Co-variates

The co-variates used in this study were trust propensity, pre-study trustworthiness in robots, prior experience with autonomous UGVs, and familiarity with autonomous UGVs. Each of these were measured as part of a pre-test questionnaire. Trust propensity was measured with an adapted 6-item instrument based on (Jessup, Schneider, Alarcon, Ryan, & Capiola, 2019). Pre-study trustworthiness in robots was measured with an adapted 9 item questionnaire based on (Esterwood & Robert, 2023) and was designed to measure subject's perceptions of robots in general and not the specific robots (i.e. UGVs) they interacted with during this study. Prior experience and familiarity with AVs were measured via single 7 point Likert scale items based on previous work (Azevedo-Sa et al., 2020). Responses to each of these measures were then used as covariates (i.e. nuisance variable) in our analysis. Multicollinearity was assessed using variance-inflation factors (VIF).

Dependent Variable

The dependent variable in this study was participants' trustworthiness in the specific robots (i.e. UGVs) they interacted with. Trustworthiness was measured as part of our post-test questionnaire. This measure was adapted from (Esterwood & Robert, 2023) and examines trustworthiness by measuring subject's perceptions of the UGVs' ability, benevolence, and integrity. The overall reliability of this measure was $\alpha = 0.76$ while the reliabilities for the ability, integrity, and benevolence sub-components were $\alpha = 0.60$, $\alpha = 0.70$, and $\alpha = 0.85$ respectively. This measure therefore possessed sufficient reliability for this preliminary analysis and results of this measure were used in our generalized linear model as the outcome variable.

Participants

In total, we recruited 20 participants for this study (10 per condition). These participants were assigned to one of our two conditions. 40% were male and the average age across participants was 35 (Std Dev = 16). Participants were recruited via online and print advertising and were compensated \$25 with the opportunity to earn bonus payments of up to \$15. The study took between 30–45 minutes to complete. This research complied with the American Psychological Association Code of Ethics and was approved by the institutional review board at the University of Michigan. Informed consent was gathered upon participants' arrival and initial on-boarding.

Procedure

Subjects were trained on the virtual environment, hardware, and GUI to be used during the study. The training included two takeover requests in a different environment from the main study session. After the training, subjects completed a pre-test survey and were reminded that UGVs may not be reliable. They were also reminded that their payment was dependent on their performance in the secondary task and how long it took the UGVs to reach their destinations. During the main study session, subjects encountered 6 takeover requests and 2 trust violations followed by either a *promise* or *no repair*. The study concluded after the 6th takeover request, and subjects completed a post-test questionnaire measuring trust.

RESULTS

To examine the impact of promises on trustworthiness we conducted a series of generalized linear models (GLM). These models used experimental condition – *no repair* or *promises* – as the predictor with trust propensity, pre-study trustworthiness in robots, prior experience with AVs, and familiarity with AVs as covariates. The predictor varied between models and was either trustworthiness overall, ability, integrity, or benevolence. This allowed us to determine what overall impact *promises* may have on trustworthiness as well as to isolate the individual effects of *promises* on ability, integrity, and benevolence. The following sections report the results of each of these models.

Promises & Trustworthiness Overall

Results predicting trustworthiness overall, as shown in Table 1, showed a marginally significant difference in trustworthiness between the *promise* condition and the *no repair* condition with an R^2 of 0.68. Furthermore, the variance inflation factors for each of the covariates were minimal ($VIF < 3$). A visual examination of estimated means via figure 1a indicates that that trustworthiness was higher for subjects in the promises condition than those in the no repair condition. The full results of this model are detailed in table 1.

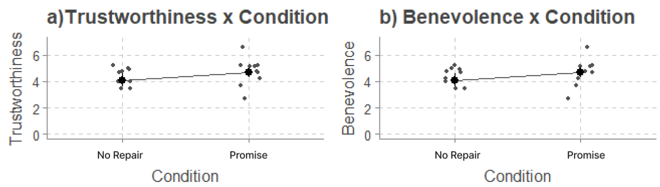

Figure 1. a)Estimated marginal means for trustworthiness in UGVs overall by experimental condition. b)Estimated marginal means for perceptions of benevolence in UGVs by experimental condition.

Table 1. *Generalized linear model appended with variance inflation factors (VIF) for predicting trustworthiness overall by condition.*

Trustworthiness in UGVs Overall				
Predictors	Estimates	CI	p	VIF
(Intercept)	1.06	-1.04 – 3.17	0.323	—
Condition [Promise]	0.59	-0.02 – 1.21	**0.059**	—
Trust Propensity	0.32	-0.21 – 0.84	0.2	1.8
AV Familiarity	0.15	-0.24 – 0.55	0.453	1.9
AV Experience	-1.01	-1.78 – -0.25	**0.009**	2.1
Pre-Trustworthiness in Robots	0.61	0.16 – 1.05	**0.007**	1.8
Observations	20			
R2	0.677			

Promises, Ability, Integrity, & Benevolence

As shown in Table 2, results predicting the subject's perceptions of ability and integrity showed no significant difference between the *promise* condition and the *no repair* condition but results predicting the subject's perceptions of benevolence did. The R^2 for these models was 0.54, 0.19, and 0.19 respectively and multicollinearity was minimal ($VIF < 3$). A visual examination of estimated means for benevolence via figure 1b indicates that for subjects in the *promise* condition, benevolence appeared to be higher.

Table 2. *Generalized linear models appended with variance inflation factors (VIF) for the models predicting perceptions of UGV ability, integrity, and benevolence by condition.*

Perceptions of Ability in UGVs				
Predictors	Estimates	CI	p	VIF
(Intercept)	5.79	2.94 – 8.63	**<0.001**	—
Condition [Promises]	0.06	-0.77 – 0.89	0.887	—
Trust Propensity	-0.27	-0.98 – 0.44	0.456	1.8
AV Familiarity	-0.01	-0.54 – 0.53	0.985	1.9
AV Experience	-1.46	-2.50 – -0.43	**0.005**	2.1
Pre-Trustworthiness in Robots	0.59	-0.01 – 1.19	0.054	1.8
Observations	20			
R2	0.543			
Perceptions of Integrity in UGVs				
Predictors	Estimates	CI	p	VIF
(Intercept)	2.20	-1.87 – 6.28	0.289	—
Condition [Promise]	0.33	-0.86 – 1.52	0.583	—
Trust Propensity	-0.00	-1.02 – 1.02	0.995	1.8
AV Familiarity	0.25	-0.52 – 1.02	0.522	1.9
AV Experience	-0.54	-2.01 – 0.94	0.478	2.1
Pre-Trustworthiness in Robots	0.55	-0.31 – 1.41	0.208	1.8
Observations	20			
R2	0.187			
Perceptions of Benevolence in UGVs				
(Intercept)	-4.81	-7.31 – -2.30	**<0.01**	—
Condition [Promise]	1.37	0.64 – 2.10	**<0.01**	—
Trust Propensity	1.23	0.60 – 1.86	**<0.01**	1.8
AV Familiarity	0.21	-0.26 – 0.68	0.387	1.9
AV Experience	-1.04	-1.94 – -0.13	**0.025**	2.1
Pre-Trustworthiness in Robots	0.68	0.15 – 1.21	**0.012**	1.8
Observations	20			
R2	0.187			

Summary of Results

Overall the results of our analysis indicate that promises appear to have a marginally significant effect on human's trustworthiness perceptions of UGVs. This effect does not appear significant for the trustworthiness sub-components of ability and integrity but does appear significant for benevolence. A closer examination of this effect based on estimated marginal means reveals that the impact on trustworthiness and benevolence is positive. We can conclude that subjects who encountered promises saw greater trustworthiness overall and perceived the UGVs as more benevolent than those who did not.

DISCUSSION & FUTURE WORK

These findings offer two contributions. First, these results examine promises which have been previously overlooked. In doing so, these findings provide new insight into what trust repairs may be effective when delivered by UGVs. In particular, these preliminary results indicate that promises may be an effective method of restoring trustworthiness – and by extension trust – after UGV failures. This is because trustworthiness overall was higher in the *promise* condition than in the *no repair* condition indicating that promises had a positive impact. This finding is especially important as other trust repair strategies (apologies, denials, and explanations) have mostly been ineffective in restoring trust (Feng & Tan, 2022; Schelble et al., 2022; Xu & Howard, 2022) and trustworthiness (Kohn et al., 2018).

The second contribution comes from our examination of not only trustworthiness overall but its different subcomponents. In particular, these preliminary results indicate that a robot's ability and integrity were not impacted by promises but benevolence was impacted. This implies that promises restore trustworthiness through increasing benevolence. This finding echoes other findings from the HRI literature more broadly (Esterwood & Robert, 2021, 2023; Lyons et al., 2023). This finding can be used to inform future theoretical development related to trust repair. Additionally, this study also adds to the literature by specifically examining the impacts of trust repair on micro factors such as benevolence. In doing so, this paper identifies a new mediating mechanism for understanding trust repair in UGVs. Future work is needed to compare other strategies.

ACKNOWLEDGEMENT

We would like to acknowledge the technical and financial support of the Automotive Research Center (ARC) in accordance with Cooperative Agreement W56HZV-19-2-0001 U.S. Army DEVCOM Ground Vehicle Systems Center (GVSC) Warren, MI.

REFERENCES

Azevedo-Sa, H., Jayaraman, S. K., Esterwood, C. T., Yang, X. J., Robert Jr, L. P., & Tilbury, D. M. (2020). Comparing the effects of false alarms and misses on humans' trust in (semi) autonomous vehicles. In *Companion of the 2020 ACM/IEEE international conference on human-robot interaction* (pp. 113–115).

Azevedo-Sa, H., Jayaraman, S. K., Esterwood, C. T., Yang, X. J., Robert Jr, L. P., & Tilbury, D. M. (2021). Real-time estimation of drivers' trust in automated driving systems. *International Journal of Social Robotics*, *13*(8), 1911–1927.

Baker, A. L., Phillips, E. K., Ullman, D., & Keebler, J. R. (2018). Toward an understanding of trust repair in human-robot interaction: Current research and future directions. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *8*(4), 1–30.

Chen, J., & Terrence, P. I. (2009). Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics*, *52*(8), 907–920.

Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A. K., Yang, X. J., & Robert Jr, L. P. (2019). Look who's talking now: Implications of av's explanations on driver's trust, av preference, anxiety and mental workload. *Transportation research part C: emerging technologies*, *104*, 428–442.

Esterwood, C., & Robert, L. P. (2021). Do you still trust me? human-robot trust repair strategies. In *2021 30th IEEE international conference on robot & human interactive communication (ro-man)* (pp. 183–188).

Esterwood, C., & Robert, L. P. (2022). A literature review of trust repair in hri. In *2022 31st IEEE international conference on robot and human interactive communication (ro-man)* (pp. 1641–1646).

Esterwood, C., & Robert, L. P. (2023). Three strikes and you are out!: The impacts of multiple human-robot trust violations and repairs on robot trustworthiness. *Computers in Human Behavior*, 107658.

Feng, Y., & Tan, H. (2022). Comfort or promise? investigating the effect of trust repair strategies of intelligent vehicle system on trust and intention to use from a perspective of social cognition. In *Cross-cultural design: 14th international conference, ccd 2022, june 26–july 1, 2022, proceedings, part iv* (pp. 154–166).

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition, Technology & Work*, *14*, 39–49.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, *53*(5), 517–527.

Jessup, S. A., Schneider, T. R., Alarcon, G. M., Ryan, T. J., & Capiola, A. (2019). The measurement of the propensity to trust automation. In *International conference on human-computer interaction* (pp. 476–489).

Johns, J. L. (1996). A concept analysis of trust. *Journal of advanced nursing*, *24*(1), 76–83.

Kim, W., Kim, N., Lyons, J. B., & Nam, C. S. (2020). Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modelling approach. *Applied ergonomics*, *85*, 103056.

Kohn, S. C., Momen, A., Wiese, E., Lee, Y.-C., & Shaw, T. H. (2019). The consequences of purposefulness and human-likeness on trust repair attempts made by self-driving vehicles. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 63, pp. 222–226).

Kohn, S. C., Quinn, D., Pak, R., De Visser, E. J., & Shaw, T. H. (2018). Trust repair strategies with self-driving vehicles: An exploratory study. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 62, pp. 1108–1112).

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied ergonomics*, *66*, 18–31.

Kramer, R. M., & Lewicki, R. J. (2010). Repairing and enhancing trust: Approaches to reducing organizational trust deficits. *Academy of Management annals*, *4*(1), 245–277.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270.

Lewicki, R. J., & Brinsfield, C. (2017). Trust repair. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 287–313.

Liu, P., Guo, Q., Ren, F., Wang, L., & Xu, Z. (2019). Willingness to pay for self-driving vehicles: Influences of demographic and psychological factors. *Transportation Research Part C: Emerging Technologies*, *100*, 306–317.

Liu, P., Yang, R., & Xu, Z. (2019). Public acceptance of fully automated driving: Effects of social trust and risk/benefit perceptions. *Risk Analysis*, *39*(2), 326–341.

Lyons, J. B., aldin Hamdan, I., & Vo, T. Q. (2023). Explanations and trust: What happens to trust when a robot partner does something unexpected? *Computers in Human Behavior*, *138*, 107473.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, *20*(3), 709–734.

Mühl, K., Strauch, C., Grabmaier, C., Reithinger, S., Huckauf, A., & Baumann, M. (2020). Get ready for being chauffeured: Passenger's preferences and trust while being driven by human and automation. *Human factors*, *62*(8), 1322–1338.

Ni, J., Hu, J., & Xiang, C. (2021). A review for design and dynamics control of unmanned ground vehicle. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, *235*(4), 1084–1100.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230–253.

Payre, W., Cestac, J., & Delhomme, P. (2016). Fully automated driving: Impact of trust and practice on manual control recovery. *Human factors*, *58*(2), 229–241.

Robert, L., & You, S. (2013). Are you satisfied yet? shared leadership, trust and individual satisfaction in virtual teams. In *Proceedings of the iconference.*

Robert, L. P., Denis, A. R., & Hung, Y.-T. C. (2009). Individual swift trust and knowledge-based trust in face-to-face and virtual team members. *Journal of Management Information Systems*, *26*(2), 241–279.

Rossi, A., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2020). How social robots influence people's trust in critical situations. In *2020 29th IEEE international conference on robot and human interactive communication (ro-man)* (pp. 1020–1025).

Rossiter, A. (2020). Bots on the ground: an impending UGV revolution in military affairs? *Small Wars & Insurgencies*, *31*(4), 851–873.

Schelble, B. G., Lopez, J., Textor, C., Zhang, R., McNeese, N. J., Pak, R., & Freeman, G. (2022). Towards ethical ai: Empirically investigating dimensions of ai ethics, trust repair, and performance in human-ai teaming. *Human Factors*, 00187208221116952.

Sebo, S. S., Krishnamurthi, P., & Scassellati, B. (2019). "i don't believe you": Investigating the effects of robot trust violation and repair. In *2019 14th ACM/IEEE international conference on human-robot interaction (hri)* (pp. 57–65).

Seet, M. S., Dragomir, A., Mathialagan, I., Ann, L. Y., binte Zaid, Z., Ramapatna, S. L., … Bezerianos, A. (2020). Subtype divergences of trust in autonomous vehicles: Towards optimisation of driver–vehicle trust management. In *2020 IEEE 23rd international conference on intelligent transportation systems (itsc)* (pp. 1–6).

Surendran, V., Mokhtari, K., & Wagner, A. R. (2021). Your robot is watching 2: Using emotion features to predict the intent to deceive. In *2021 30th IEEE international conference on robot & human interactive communication (ro-man)* (pp. 447–453).

Xu, J., & Howard, A. (2020). Would you take advice from a robot? Developing a framework for inferring human-robot trust in time-sensitive scenarios. In *2020 29th IEEE international conference on robot and human interactive communication (ro-man)* (pp. 814–820).

Xu, J., & Howard, A. (2022). Evaluating the impact of emotional apology on human-robot trust. In *2022 31st IEEE international conference on robot and human interactive communication (ro-man)* (pp. 1655–1661).

Ye, S., Neville, G., Schrum, M., Gombolay, M., Chernova, S., & Howard, A. (2019). Human trust after robot mistakes: Study of the effects of different forms of robot communication. In *2019 28th IEEE international conference on robot and human interactive communication (ro-man)* (pp. 1–7).

Yokoi, R., & Nakayachi, K. (2021). Trust in autonomous cars: exploring the role of shared moral values, reasoning, and emotion in safety-critical decisions. *Human factors*, *63*(8), 1465–1484.

You, S., & Robert Jr, L. P. (2018). Human-robot similarity and willingness to work with a robotic co-worker. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 251–260).

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation research part C: emerging technologies*, *98*, 207–220.