

Performance and Relative Risk Dynamics during Driving Simulation Tasks under Distinct Automation Conditions

Lucero Rodriguez Rodriguez ^a, Carlos Bustamante Orellana ^a, Gregory M. Gremillion ^b, Lixiao Huang ^a, Mustafa Demir ^a, Nancy Cooke ^a, Jason S. Metcalfe ^b, Polemnia G. Amazeen ^a, Yun Kang ^a

^aArizona State University, ^bArmy Research Laboratory

Risk has been a key factor influencing trust in Human-Automation interactions, though there is no unified tool to study its dynamics. We provide a framework for defining and assessing relative risk of automation usage through performance dynamics and apply this framework to a dataset from a previous study. Our approach allows us to explore how operators' ability and different automation conditions impact the performance and relative risk dynamics. Our results on performance dynamics show that, on average, operators perform better (1) using automation that is more reliable and (2) using partial automation (more workload) than full automation (less workload). Our analysis of relative risk dynamics indicates that automation with higher reliability has higher relative risk dynamics. This suggests that operators are willing to take more risk for automation with higher reliability. Additionally, when the reliability of automation is lower, operators adapt their behavior to result in lower risk.

INTRODUCTION

Human-automation interaction research is critical to improve human-machine teaming performance, and dynamic measures of the interactions are the key. The term *trust in automation* emerged through this research and has been widely studied through self-assessed questionnaires (Itoh & Tanaka, 2000; J. Lee & Moray, 1992; Xu & Dudek, 2015). In the literature there exist variations on the definition of *trust*. Kramer (1999, p. 571) defined it as "a state of perceived vulnerability or risk that is derived from an individual's uncertainty regarding the motives, intentions, and perspective actions of others on whom they depend". Mayer et al. (1995, p. 712) definition is widely accepted due to its relevance to Human-Human interaction and Human-Automation interaction (J. D. Lee & See, 2004; Stuck, 2020). They define trust as a "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that party". Both definitions link the terms *vulnerability* and *risk* to trust and identify that individuals must be willing to put themselves at risk.

Risk is defined as "the product of the probability of an event and the consequences that necessarily accompany that event" (Sheridan, 2008, p. 418). Others define it as "the extent to which there is uncertainty about whether potentially significant and/or disappointing outcomes of decisions will be realized" (Sitkin & Pablo, 1992, p. 10). Research indicates that risk is a key factor influencing trust in the Human-Automation interaction (Chancey et al., 2017; Hoff & Bashir, 2015; Kramer, 1999; Loft et al., 2021; Lyons & Stokes, 2012; Mayer et al., 1995; Sato et al., 2020). Lyons and Stokes (2012) reported an increased reliance on automation when operators were faced with higher-risk decisions. Chancey et al. (2017) observed an effect of risk on task performance. Sato et al. (2020) found that a high risk environment increased trust with high workload conditions. Conversely, Loft et al. (2021) reported that participants were more likely to reject automated advice for high-risk decisions. Although trust has repeatedly been conceptualized as having a component of risk, the role risk plays, as well as what elements

of risk impact trust (e.g., perceived risk, risk-taking propensity), has not been clearly defined and studied (Stuck et al., 2021).

Research suggests that self-confidence which can be reflected through the operator's performance, is an important factor that guides trust formation (Hoff & Bashir, 2015; J. D. Lee & See, 2004). Gao and Lee (2006) used decision field theory to mathematically model the preference of automation usage as a difference of trust in automation and self-confidence. They define trust in automation as a function of the automation's performance and self-confidence as a function of the operator's performance. In addition, de Vries et al. (2003), showed that the difference between the measure of trust and self-confidence is highly predictive of automation reliance decision-making.

Motivated by Gao and Lee (2006) and Stuck et al. (2021) work, we provide a framework for defining a time dependent measure of relative risk of automation usage decision-making. This measure is the ratio between the risk of automation usage and the risk of manual control, which is measured by the automation's and operator's performance. Thus, this measure can aid the operator in the decision-making process of engaging or not the automation. The current study addresses the following research questions: (1) what are the operator's performance dynamics across different types and levels of automation? (2) what are the relative dynamics across different types and levels of automation? and (3) what are the relative dynamics on participants with different level of performance?

METHODS

Experiment Design

Testbed. Sixteen participants completed a "leader-follower" task with the assistance of an automation driving system in a simulated environment. Additional information about participants, task descriptions, measures and analyses can be found in Drnec and Metcalfe (2016), Gremillion et al. (2016), Neubauer et al. (2020) and Rodriguez et al. (2021). The following information only highlight the key components involved in calculating the relative risk dynamics.

Condition. The experiment had a manual condition and a two (automation type: speed, full) by two (automation level: low, high) within subjects design. The speed automation included cruise control, and full automation included cruise control and lane-keeping. The two levels of automation correspond to the reliability level of the automation (e.g., low reliability has a higher variance in the maintenance of distance/lane position). Hence, the testbed contained five Automation Conditions (AC): Manual (*M*), Speed High (*SH*), Speed Low (*SL*), Full High (*FH*), and Full Low (*FL*).

Procedure. Participants completed the manual condition as a practice lap to acquaint themselves with the driving simulator. For the remaining four conditions, participants were informed of the type of automation, but not the level of automation and the mean lap time was approximately 12 minutes. Participants were monetarily compensated for their participation and overall driving performance score. All five conditions were scored, but the manual condition did not count against their monetary compensation.

Tasks and Scoring. Participants are required to maintain the car's lane position, 5-20 meters behind a lead vehicle, and monitor for pedestrians on the simulation. These tasks were enforced by the introduction of three types of perturbations. The first perturbation was a change in the lead vehicle's velocity, which required accelerating or decelerating to maintain the required distance and avoid a collision. The second perturbation was wind gusts that affected the participant's vehicle lane position and required to adjust steering. The third perturbation was the appearance of pedestrians, which could stand on the side of the road, or walk across the car's path. This required proper identification of pedestrian type via button press, which triggered their disappearance from the simulation.

Participants began each trial with 500 pts and could see their changing score in the display. Points were deducted for task violations such as deviations from the lane and distance from the lead vehicle (-2 pts); incorrect or missed button presses (-5 pts); and collisions with vehicles (-50 pts) or pedestrians (-100 pts). Additionally, participants were awarded 100 pts for completion of the driving course, which were awarded evenly through the course. Thus, participants' possible maximum final score was 600 pts.

Operator's Performance Classification

Performance of each operator can be reflected and measured by their score. Literature suggests that automation condition and operator ability can impact the operators' performance (Hoff & Bashir, 2015). The performance of operator i at time t under certain ACs (M , SL , SH , FL , FH) is denoted by $P_i^{AC}(t)$. With these terms, we have the mean performance dynamics of all 16 operators over certain automation condition $\bar{P}^{AC}(t) = \frac{\sum_{i=1}^{16} P_i^{AC}(t)}{16}$. Thus we can explore how automation type (speed, full) and automation level (low, high) impact the mean performance dynamics by comparing dynamics of $\bar{P}^M(t)$, $\bar{P}^{SL}(t)$, $\bar{P}^{SH}(t)$, $\bar{P}^{FL}(t)$ and $\bar{P}^{FH}(t)$.

Each operator i has five performance dynamics $P_i^{AC}(t)$ due to the five AC, thus each operator has five final scores $P_i^{AC}(f)$.

The ability of each operator is reflected by the final scores of the five ACs. We say operator i belongs to *Low-Performing Group* if all five final scores are below 300 (e.g., operator 4 and 13) and operator i belongs to *High-Performing Group* if all five final scores are above 400 (e.g., operator 2, 5 and 12). The remaining eleven operators belong to *Medium-Performing Group* whose five final scores range from 0 to 600, see Figure 1.

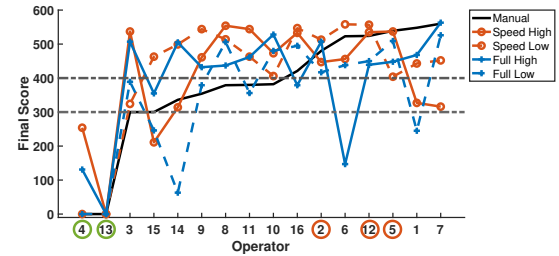


Figure 1. Operator's final scores. The figure helps classify the operator's performance based on their final scores over five automation conditions (M , SH , SL , FH , FL). Operators are ordered in ascending order based on their final score in condition M .

Relative Risk

The concept of risk in automation usage has not been clearly defined and measured. Our goal is to provide a comparison measure between the risk of automation usage and the risk of manual control. We consider that the experimental design only registered penalties due to task violations and not awards for task completions. Thus, we measure risk of automation usage through the points lost with automation usage and risk of manual control through the points lost when automation was disengaged. In epidemiology (Schmidt & Kohlmann, 2008; Zhang & Yu, 1998), the risk ratio or Relative Risk (RR) has been one of the most widely used measures of association in disease spreading. Hence, we can define the *relative risk* of automation usage as the comparison of the risk of using the automation with the risk of not using the automation. In our experiment, we define the relative risk dynamics of operator i under certain automation condition AC_{rr} (SL , SH , FL , FH) as follows:

$$RR_i^{AC_{rr}}(t) = \frac{\text{Risk of automation usage}}{\text{Risk of manual control}} = \frac{\text{Total Lost Points with automation on over } [0, t]}{\text{Total Lost Points without automation over } [0, t] + 1} \quad (1)$$

where the 1 in the denominator guarantees the meaningfulness of the equation when the operators have not lost points while having manual control. For example, the notation of $RR_i^{SL}(t)$ is calculated as the ratio between the total lost points of operator i during Speed Low automation over time $[0, t]$ and the total lost points of operator i without automation over time $[0, t]$ plus one (see Eq. 1).

Per definition of RR, we expect its dynamics lay in the range of $[0, 600]$ where $RR_i^{AC_{rr}}(t) = 0$ if there are no points lost when automation is on, and $RR_i^{AC_{rr}}(t) = 600$ if all possible 600 pts are lost with automation usage. This property is inherited to the calculation of the average relative risk, $\bar{RR}(t)$, of a certain group and automation condition. Our dynamic RR measure accounts from the start of the trial. This is important since participants could observe their score; thus, it im-

pacts the perception of risk of automation usage from their experience on the trial. We could expect to use relative risk dynamics to aid decision-making of automation usage to achieve the best outcomes (e.g., better score/performance). A value of $RR_i^{ACrr}(t) > 1$ indicates a higher probability of losing points while using automation. Thus, the rational decision would be to use manual control to maintain good performance. In the following section, we first explore how automation type (speed, full) and automation level (low, high) impact the mean RR dynamics of all participants by comparing dynamics of $\overline{RR}^{SL}(t)$, $\overline{RR}^{SH}(t)$, $\overline{RR}^{FL}(t)$ and $\overline{RR}^{FH}(t)$. Second, we explore the impact of operator's ability on their RR dynamics over four automation conditions $\overline{RR}_a^{ACrr}$ where a corresponds to Low-(L), Medium-(M), and High-Performing(H) Group, respectively.

RESULTS

In this section we do an exploratory study of RR dynamics on the four different types of automation. For convenience, our figures across varied automation conditions use the colors orange (speed automation) and blue (full automation), solid or dash line for high and low automation level, respectively, and black solid line for manual condition. In figures focusing on varied level of performance groups, we use green, blue, and orange for Low, Medium, and High-Performing Group, respectively.

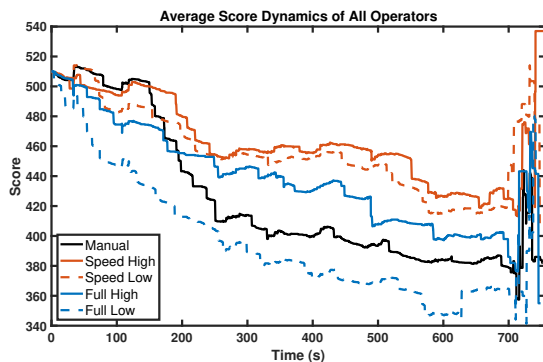


Figure 2. The figure shows the mean performance dynamics of all participants across five automation conditions.

Performance Dynamics Across Automation Conditions

Figure 2 shows the average performance dynamics of all 16 operators over the five experimental conditions. The manual condition performs best during the initial 200 seconds. During the interval from 200 to 700 sec, the order of the average performance dynamics is $SH > SL > FH > M > FL$. Thus, on average, operators are performing better (1) in speed automation (more workload) than full automation (less workload); and (2) in high automation than low automation reliability. On average the $\overline{RR}$ value in FH is 48.8 ($\sigma = 8.5$) higher than under FL automation; and for SH is 8.2 ($\sigma = 4.7$) higher than under SL automation. Additionally, the high-reliability automation was able to perform its tasks with more precision, in comparison to the low reliability automation. Thus, operators performed better with a high-level automation in comparison to the low-level automation.

Relative Risk Dynamics in varied Automation Conditions

Figure 3 depicts the average RR dynamics of all 16 participants over four AC. Although performance (Figure 2) and RR dynamics are associated, RR dynamics offers a detailed analysis of the operator's behavior and observation of a decreased performance during automation usage ($\nearrow RR$) or manual control ($\searrow RR$) at any time t . For example, we observe that (1) during the first 250 sec of the trial SL and FL automation conditions have a quick increase of RR, where $\overline{RR}(t)$ for FL automation is larger than SL . Thus, participants lost a considerable number of points while using the automation during the first 250 sec. Conversely, we observe an opposite behavior for SH and FH as the RR increases steadily. Moreover, (2) the $\overline{RR}(t)$ for FH automation is similar or slightly larger than SH automation. After the initial 250 sec of the trial, (3) RR in high automation reliability (solid lines) is larger than in low automation reliability (dashed lines). While $\overline{RR}(t)$ becomes larger in FH than in SH , we observe an opposite event for the low automation reliability. That is, (4) $\overline{RR}(t)$ becomes smaller for FL automation in comparison to SL automation.

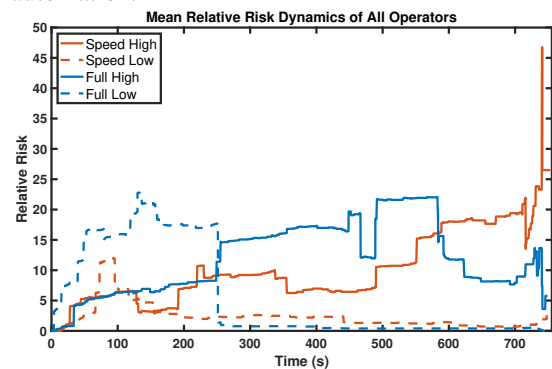


Figure 3. Average relative risk dynamics of all sixteen participants across SL , SH , FL , and FH automation conditions.

Impact of Operator's Performance Level on Relative Risk Dynamics Across Varied Automation Conditions

Figure 4 depicts the average RR dynamics of Low-, Medium-, and High-Performing Groups across the four different automation conditions. The results show that (1) the range of the average relative risk ($\overline{RR}(t)$) dynamics with low automation reliability (dashed lines, Fig. 4b and 4d) are larger than with high automation reliability (solid lines, Fig. 4a and 4c). This is due to the high increase of RR values in the Low-Performing Group occurring around 100 sec. Additionally, we (2) observe a distinctive "hump" type of RR dynamics for this group (green line). This dynamic occurs in the first 200-250 sec with the peak value occurring around 100 sec. The behavior is present only for SH , SL , and FL automation conditions with an average of 15.4 ($\sigma = 13.26$), 16.4 ($\sigma = 18.8$), and 85 ($\sigma = 42.1$) higher, correspondingly, in comparison to Medium-Performing Group. Moreover, (3) the $\overline{RR}(t)$ values for SH , SL and FL conditions are ordered as Medium>High>Low-Performing Group after a certain time for each condition. Furthermore, (4) Medium- and High-Performing Groups have smaller $\overline{RR}(t)$ dynamics during low automation reliability and Medium-Performing Group has higher RR value than the High-Performing Group. In addition

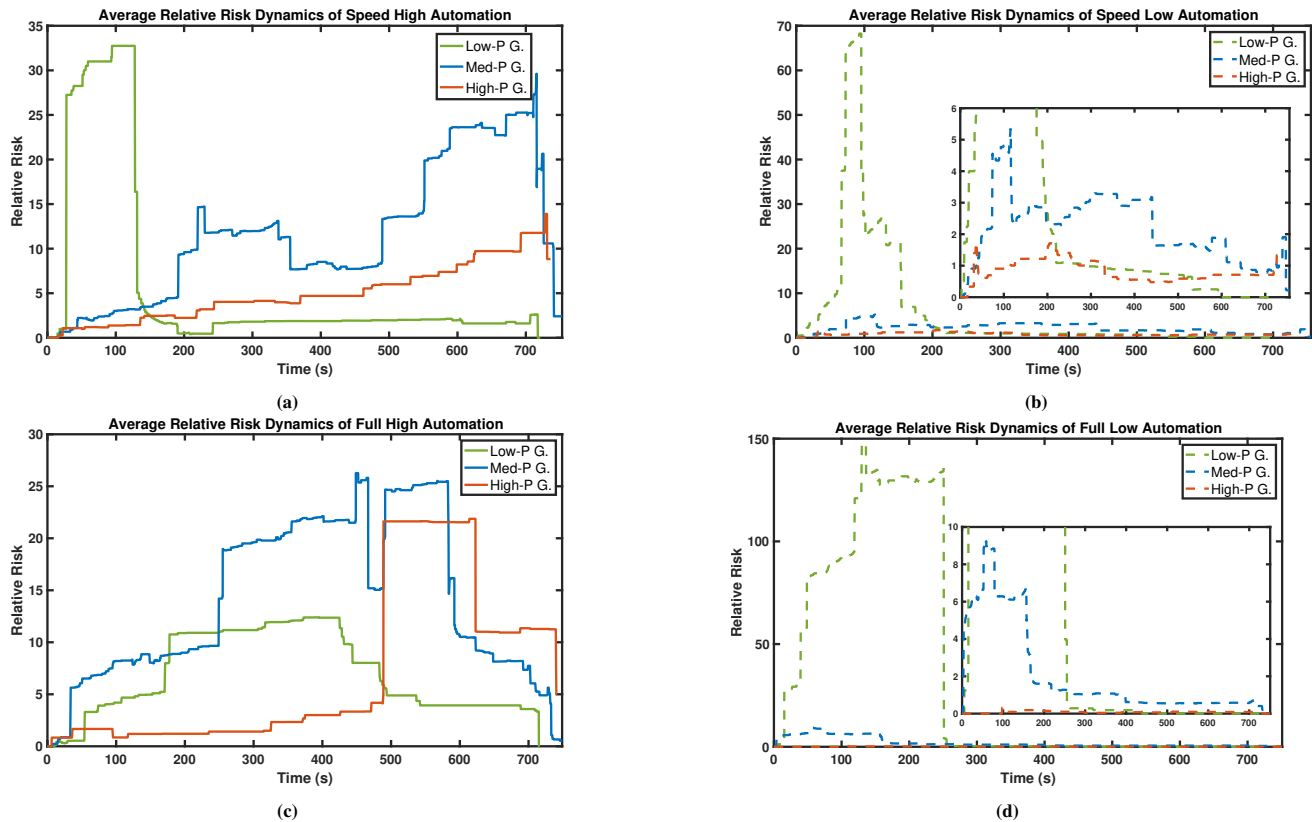


Figure 4. The figures show the average relative risk dynamics of Low, Medium, High-Performing Group across four different automation conditions.

to this, during *FL* automation, all performance groups maintain a RR below the value of one. We identified specific dynamics pertaining to the operators' level of performance and type of automation. The (5) Low-Performing Group has a large spike of relative risk at the beginning of trial *SH*, *SL* and *FL*. The (6) Medium-Performing Group shows two distinct types of dynamics. The first type, observed in trials *SL* and *FL*, resembles the Low-Performing group with considerably smaller RR values. The second type, observed in trials *SH* and *FH*, resembles an oscillatory step function. The (7) High-Performing Group displays three types of RR dynamics. The first type, observed in trials *SH* and *FH*, is an increasing function; the second, observed in *SL*, is an oscillatory function, and the third, observed in *FL*, is a constant function.

DISCUSSION

In this study, we defined and derived a time-dependent measure to compare the risk of automation usage and risk of manual control to assist decision-making of automation usage. This measure was derived through the operators' performance dynamics, which allowed us to identify that their performance was better when using partial automation than when using full automation. This could be due to the lower number of tasks that need to be performed by the automation when it is being used. Additionally, a higher workload might maintain the participant engaged and alert to all the required tasks; that is, while using the full automation the operators might experience a sense of ease and lose concentration on the required task. Furthermore, performance dynamics demonstrated that operators always per-

form better using automation with high reliability.

The analysis of relative risk dynamics reflects that (1) during partial automation (i.e., more workload) and full automation (i.e., less workload) with low reliability, operators used automation soon after starting the simulation (before 250 sec) and were not able to assess the usage of a low reliable automation. Therefore, (2) operators lost more points when using automation at the beginning of the trial, which is represented by a spike in relative risk. These results could reflect the trust calibration process and the development of learned trust. In this initial process, the operator who encounters a novel automation gathers experiential data of the automation performing the task to inform their level of trust in that system (Hoff and Bashir, 2015). During the early stage of the trial, the operator is likely to overtrust the automation, expecting its capability to be greater than it truly is. Hence, the operator inadvertently allows the automation more frequent opportunities to fail before appropriately calibrating their level of trust and reducing reliance on automation. Moreover, (3) partial and full automation with high reliability display dynamics resembling an increasing step function. Then the initial trust calibration, and formation of learned trust results from few failures due to high-reliability automation. Hence, the user may increase their reliance on the automation over time. Additionally, (4) operators used the automation more towards the middle and end of the drive, which indicates they learned about the automation reliability during the first 250 sec of each trial. The study of the dynamics pertaining to the operator's level of performance reflects (5) a large spike of relative risk at the beginning of the trials in the Low-Performing Group. This result may suggest that operators' decisions of when to

use the automation are particularly poor and represent poorly calibrated situational trust, as they use the automation in situations where it underperforms and disuse it when it would be most beneficial. Instead, the High-Performing Group demonstrates across conditions behaviors that are indicative of much more appropriate levels of trust, where RR gradually increases as they rely increasingly on a high-performing automation. We hypothesize that this behavior contributes to the three types of RR dynamics for this group. Finally, we observe that (6) the maximum value of $\overline{RR}(t)$ drastically increases from an automation with high-reliability to a low-reliability automation in the Low-Performing Group.

The contributions of this work are (i) the study of operators' performance dynamics across different types and levels of automation; (ii) a precise definition and mathematical derivation of the measurable variable *relative risk* for decision-making in automation usage; (iii) the study of relative risk dynamics across different types and levels of automation; and (iv) the study of relative risk dynamics on operators with different level of performance. This framework is applicable to other testbeds that record the operator's and automation's performance through a scoring system and can take into account a reward system for successfully completed tasks. A limitation to this framework is the inability to assess complex risks (e.g., risk of losing a human life).

ACKNOWLEDGEMENTS

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-20-2-0252, the Research Assistantship awarded by the School of Human Evolution and Social Change from Arizona State University, and the James S. McDonnell Foundation 21st Century Science Initiative in Studying Complex Systems Scholar Award (UHC Scholar Award 220020472). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

Chancey, E. T., Bliss, J. P., Yamani, Y., & Handley, H. A. H. (2017). Trust and the compliance-reliance paradigm: The effects of risk, error bias, and reliability on trust and dependence [PMID: 28430544]. *Human Factors*, 59(3), 333–345. <https://doi.org/10.1177/0018720816682648>

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning [Trust and Technology]. *International Journal of Human-Computer Studies*, 58(6), 719–735. [https://doi.org/10.1016/S1071-5819\(03\)00039-9](https://doi.org/10.1016/S1071-5819(03)00039-9)

Drnec, K., & Metcalfe, J. S. (2016). Paradigm development for identifying and validating indicators of trust in automation in the operational environment of human automation integration. In D. D. Schmorow & C. M. Fidopiastis (Eds.), *Foundations of augmented cognition: Neuroergonomics and operational neuroscience* (pp. 157–167). Springer International Publishing.

Gao, J., & Lee, J. (2006). Extending the decision field theory to model operators' reliance on automation in supervisory control situations. *IEEE*

Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 36(5), 943–959. <https://doi.org/10.1109/TSMCA.2005.855783>

Gremillion, G. M., Metcalfe, J. S., Marathe, A. R., Paul, V. J., Christensen, J., Drnec, K., Haynes, B., & Atwater, C. (2016). Analysis of trust in autonomy for convoy operations. In T. George, A. K. Dutta, & M. S. Islam (Eds.), *Micro- and nanotechnology sensors, systems, and applications viii* (pp. 356–365). SPIE. <https://doi.org/10.1117/12.2224009>

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3), 407–434.

Itoh, M., & Tanaka, K. (2000). Mathematical modeling of trust in automation: Trust, distrust, and mistrust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(1), 9–12. <https://doi.org/10.1177/154193120004400103>

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual review of psychology*, 50, 569–98.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243–1270. <https://doi.org/10.1080/00140139208967392>

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance [PMID: 15151155]. *Human Factors*, 46(1), 50–80. <https://doi.org/10.1518/hfes.46.1.50.30392>

Loft, S., Bhaskara, A., Lock, B. A., Skinner, M., Brooks, J., Li, R., & Bell, J. (2021). The impact of transparency and decision risk on human-automation teaming outcomes. *Human Factors*, 00187208211033445.

Lyons, J. B., & Stokes, C. K. (2012). Human-human reliance in the context of automation [PMID: 22409106]. *Human Factors*, 54(1), 112–121. <https://doi.org/10.1177/0018720811427034>

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, 20(3), 709–734. <http://www.jstor.org/stable/258792>

Neubauer, C., Gremillion, G., Perelman, B. S., La Fleur, C., Metcalfe, J. S., & Schaefer, K. E. (2020). Analysis of facial expressions explain affective state and trust-based decisions during interaction with autonomy. In T. Ahrum, W. Karwowski, A. Vergnano, F. Leali, & R. Taiar (Eds.), *Intelligent human systems integration 2020* (pp. 999–1006). Springer International Publishing.

Rodriguez, L. R., Orellana, C. B., Landfair, J., Magaldino, C., Demir, M., Amazeen, P. G., Metcalfe, J. S., Huang, L., & Kang, Y. (2021). Dynamics of trust in automation and interactive decision making during driving simulation tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 65(1), 786–790. <https://doi.org/10.1177/1071181321651288>

Sato, T., Yamani, Y., Liechty, M., & Chancey, E. T. (2020). Automation trust increases under high-workload multitasking scenarios involving risk. *Cognition, Technology & Work*, 22(2), 399–407.

Schmidt, C. O., & Kohlmann, T. (2008). When to use the odds ratio or the relative risk? *International journal of public health*, 53(3), 165.

Sheridan, T. B. (2008). Risk, human error, and system resilience: Fundamental ideas [PMID: 18689048]. *Human Factors*, 50(3), 418–426. <https://doi.org/10.1518/001872008X250773>

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *The Academy of Management Review*, 17(1), 9–38. <https://www.jstor.org/stable/258646>

Stuck, R. E. (2020). *Perceived relational risk and perceived situational risk: Scale development* (Doctoral dissertation). Georgia Institute of Technology.

Stuck, R. E., Tomlinson, B. J., & Walker, B. N. (2021). The importance of incorporating risk into human-automation trust. *Theoretical Issues in Ergonomics Science*, 0(0), 1–17. <https://doi.org/10.1080/1463922X.2021.1975170>

Xu, A., & Dudek, G. (2015). Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 221–228.

Zhang, J., & Yu, K. F. (1998). What's the Relative Risk? A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes. *JAMA*, 280(19), 1690–1691. <https://doi.org/10.1001/jama.280.19.1690>