The Impact of Automation Conditions on Reliance Dynamics and Decision-Making

Carlos Bustamante Orellana [a], Lucero Rodriguez Rodriguez [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]

[a]Arizona State University, [b]Army Research Laboratory

The decision process of engaging or disengaging automation has been termed reliance on automation, and it has been widely analyzed as a summary measure of automation usage rather than a dynamic measure. We provide a framework for defining temporal reliance dynamics and apply it to a data-set from a previous study. Our findings show that (1) the higher the reliability of an automated system, the larger the reliance over time; and (2) more workload created by the automation type does not significantly affect the operators' reliance dynamics in high-reliability systems, but it does produce greater reliance in low-reliability systems. Furthermore, on average, operators with low performance make fewer decision changes and prefer to stick to their decision of using automation even if it is not performing well. Operators with high performance, on average, have a higher frequency of decision change, and therefore, their automation usage periods are shorter.

INTRODUCTION

In recent years, human-automation interactions have become more common due to technological advancements and the necessity of reducing human workload in certain labors. Self-driving cars are one of the most known examples involving these types of interactions, with many companies working hard to achieve fully autonomous cars (Shladover, 2021). Currently, most automated systems still require human supervision to achieve better results and increase safety. Humans choose when to lend control to the automated system and when to return to manual operation (Chiou & Lee, 2021). The fraction of time that human operators have the automation engaged is usually denoted, or understood, as reliance on automation (Dzindolet et al., 2003; Lee & Moray, 1992). Inappropriate reliance on automated systems may lead to situations in which human operators depend upon a system to perform in ways for which it was not intended (misuse) or simply reject the capabilities of the system (disuse) they can use (Lee & Moray, 1992; Parasuraman, 1997). Dzindolet et al. (2003) indicated that inappropriate reliance is likely to result when operators trust an automated system that performs worst than the manual operation or distrust a system that performs better than manual control. Therefore, it is important to understand the factors that guide reliance on automation to avoid such situations.

Several research works have studied reliance as a global measure (Ezer et al., 2007; Gao & Lee, 2006; Gremillion et al., 2016; Guznov et al., 2016; Lee & Moray, 1992; Ross et al., 2008; Wang et al., 2009), i.e., they usually explore how different conditions may influence reliance at the end of a trial or experiment. For example, Lee and Moray (1992) explored how the level of fault of automated controllers influences the overall reliance of human operators, and Gremillion et al. (2016) explored how the type of automation and its reliability level impacted the overall reliance on automation. However, it is also important to analyze the dynamics of reliance on automation, i.e., how reliance changes over time due to different circumstances. Gremillion et al. (2016) and Rodriguez et al. (2021) concluded that the reliability level of the automation can substantially impact the reliance of human operators. Naturally, the human perception of the automation reliability occurs and changes dynamically while the human is using the automation. The analysis of reliance post hoc, i.e., after the use of automation, may result in the loss of information that can only be seen in a temporal dynamical analysis. Given that inappropriate reliance can negatively impact the performance of human operators, it is also important to investigate the behaviors of operators with different levels of performance.

The objective of this research is to analyze the impact of the automation type and its reliability level on the dynamics of reliance and understand how the manner in which humans rely on automation affects their performance. The present study addresses the following research questions: (1) what are the behaviors of operators with different levels of performance across different automation conditions?; (2) what are the reliance dynamics across different automation conditions?; and (3) what are the reliance dynamics on operators with different levels of performance? First, we describe the experimental design and the operators' performance classification. Then, we provide a mathematical definition for reliance on automation. Next, we analyze operators' behaviors and the dynamics of reliance. Finally, we discuss our results and future research.

METHODS

Experimental Design

Data were collected using a driving simulator at the US DEVCOM Ground Vehicle Systems Center (Drnec & Metcalfe, 2016; Gremillion et al., 2016; Neubauer et al., 2020; Rodriguez et al., 2021). A full description of the experiment and data recording can be found in the cited articles, here we highlight the critical components that are relevant to the analysis of reliance dynamics. Sixteen operators, with a minimum driving experience of 2 years, were selected for this experiment. The selected operators were monetarily compensated for their participation aside from an extra reward based on performance. Operators could see their performance, measured through a score, on the screen. They had to follow a lead car while keeping their vehicle in the correct lane and a safe distance from the lead car. They could use the assistance of an autonomous driving system at any time during their driving task. Initially, they did a training trial in which no automation was involved (Manual) to get familiarized with the driving simulator. The experiment had a two (automation type: Speed, Full) by two (automation level: Low, High) within-subjects design. Speed automation included cruise control only, whereas the Full automation included a cruise control and lane-keeping. The two levels of automation corresponded to reliability level; more reliable automation exhibited better performance in the driving tasks. Automation failures could include delays in returning to the correct distance behind the lead car and to the correct lane after a perturbation occurred. These failures result in point deductions, and therefore, lower performance. Operators performed five trials, corresponding to five automation conditions: Manual (M), Speed Low (SL), Speed High (SH), Full Low (FL), and Full High (FH).

Perturbations were imposed into the system during the experiments to require the adjustment of either the speed or steering of the vehicle. These perturbations included changes in the velocity of the lead vehicle, wind gusts, traffic vehicles, and the sudden appearance of pedestrians. Pedestrians were either static, standing on the side of the road, or dynamic, walking into the street and potentially across the operator's path. This design was employed to increase subjective cognitive workload.

Operators' Classification into Performance Groups

The performance of the operators was measured through a score. Operators started each trial with 500 points which could decrease because of task violations. These violations included but were not limited to deviations from the right lane or distance to the lead car (-2 points) and collisions with other vehicles (-50 points) or pedestrians (-100 points). Additionally, operators were awarded a total of 100 points, which were evenly distributed along nine zones, for completing the driving circuit (Gremillion et al., 2016).

The final score of each operator over five trials (M, SL, SH, FL, FH) was used to classify them into: Low-Performing Group, Medium-Performing Group, and High-Performing Group.

1. *Low-Performing Group:* operators' final scores in all five trials are below 300 points. There were two operators in this group.
2. *High-Performing Group:* operator's final scores in all five trials are above 400 points. There were three operators in this group.
3. *Medium-Performing Group:* operators who are not in the previous two groups; their final scores range between 0 and 600 points. There were 11 operators in this group.

Reliance Dynamics of Using Automation

Reliance on automation is usually denoted as the fraction of time that human operators have the automation engaged (Lee & Moray, 1992). In this paper, we extend this concept to temporal dynamics that reflect the decision-making of using or not using automation over time. Assume that operator i is performing some work over a period of time T with an option of using automation. Over the time interval $[0,t] \subset [0,T]$, assume that there are K time sub-intervals that the operator decides to use automation, i.e. $[t_1,t_2],[t_3,t_4],...,[t_K,t_{K+1}]$ with $t_{\geq}0$ and $t_{K+1} \leq t$. Then the reliance of this operator i at time $t \in [0,T]$ is defined as follows

$$R_i(t) = \frac{1}{t}\sum_{j=1}^{K}\left(t_{j+1} - t_j\right) \quad (1)$$

which indicates that $R_i(t) \in [0,1]$ for all operator i and time $t \leq T$. Reliance dynamics $R_i(t)$ of operator i at time t is the result of the operator's decision to use the automation and the time length that such operator has used it in the past.

To continue our work, we extend the concept of reliance dynamics of an individual to a group level. More specifically, we define $\bar{R}^{ac}_{group}(t)$ as the mean reliance dynamics of a certain group under a specific automation condition, where *ac* refers to the automation condition (e.g., SL, SH, FL, FH), and *group* refers to the performance level of the operators (e.g.,Low, Medium, and High-Performing Group). For example, $\bar{R}^{FH}_{Low}(t)$ refers to the mean reliance dynamics of the Low-Performing group under FH condition. We can also consider a group that contains all operators for more general analysis.

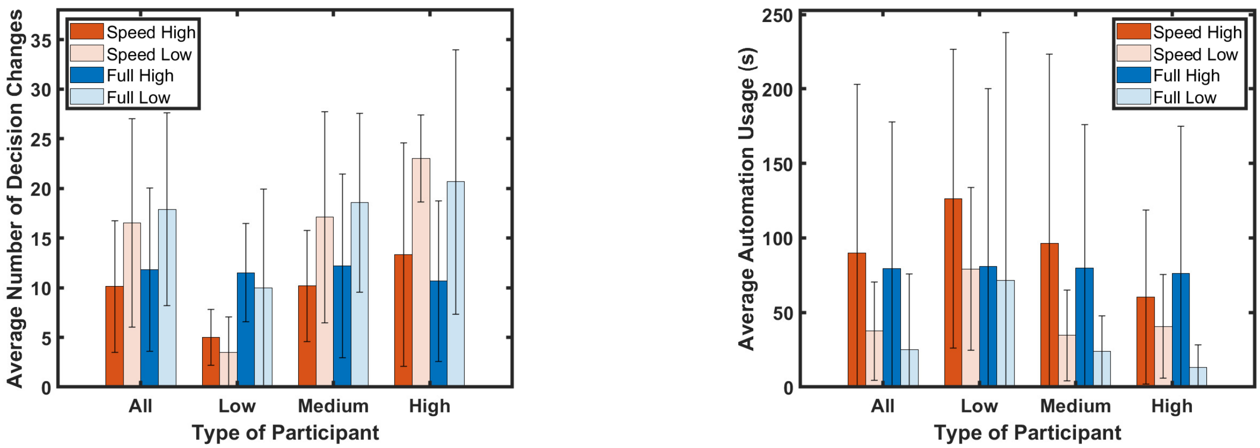
RESULTS

Based on the given reliance definition and described data, we aim to provide an analysis of how factors such as the type of automation (Speed, Full) and its reliability level (Low, High) affect the dynamics of reliance on automation for different performance groups, including the overall case. Furthermore, we assess the impact of reliance choices on the human operators' performance, across different automation conditions, through the study of the average automation usage each performance group has each time the automation is turned on, and the average number of decision changes that occur in these different groups.

In our figures, we use orange and blue colors to represent the Speed and Full types of automation, respectively; and solid and dash line styles for high and low automation reliability, respectively. For figures focusing on reliance dynamics of different performance groups, we use green, blue, and orange colors to depict Low, Medium, and High-Performing groups, respectively.

Impact of Reliance Choices on Operators' Performance

Figure 1a shows that the High-Performing Group has a higher frequency of switching the automation on-and-off in the Speed High (SH), Speed Low (SL), and Full Low (FL) conditions than that of the Low-Performing Group. We found that these observations are statistically significant using a one-tailed t-test with an alpha level of .05. We obtained $t(3) = 3.67, p = .02$; $t(3) = 10.87, p = .001$; and $t(3) = 3.48, p = .02$ for the SH, SL, and FL conditions, respectively. In addition, Medium and High-Performing groups seem to have a higher frequency of switching on-and-off when the automation reliability is low. Figure 1b depicts that the High-Performing Group has a shorter average automation duration between one switch on-and-off in the SH, SL, and FL conditions than the Low-Performing Group. We found that these observations are statistically significant using the same test as before. We obtained $t(3) = 7.91, p = .002$; $t(3) = 6.17, p = .004$; and $t(3) = 6.23, p = .004$ for the SH, SL, and FL conditions, respectively.



(a) The impact of decision making frequency on the operators' performance.

(b) Average duration of automation usage for each switch on.

Figure 1. Impacts of automation conditions and operators' reliability on decision making behavior.

The average decision-making behavior shown in Figure 1 suggests that the Medium-Performing Group, with about 69% of the population (11/16), behaves more consistently regarding the decision of using automation across all automation conditions. It is important to remark that the means shown in this figure have a large variance, especially in the Low and High-Performing groups due to the small number of operators that fall into these groups. The high variance observed in the Medium-Performing group can be due to their scores ranging from 0 to 600; a heterogeneous group.

Reliance Dynamics Over Varied Automation Conditions

Automation types and their reliability levels can contribute greatly to operators' decisions of using automation and the length of usage each time (Gremillion et al., 2016). Figure 2 depicts the average reliance dynamics of all operators over four automation conditions. Our findings indicate that (1) in high-reliability automation systems (solid lines), reliance has a rapid increase over time until around 50 seconds. Then, reliance under Speed automation seems to be stable with slight increases over time, while under Full automation, reliance decreases towards the end. This suggests that workload due to automation type of Speed versus Full may not have a significant impact on reliance dynamics. Additionally, (2) reliance in low-reliability automation systems (dashed lines), has a rapid increase over time until around 50 seconds, after which it exhibits a a steady decrease. Starting from around 50 seconds, reliance under Speed automation is constantly larger (0.11 on average) over time than reliance under Full automation. This suggests that a larger workload due to automation type (Speed workload should be larger than Full workload) can lead to higher automation usage.

There seems to be a critical time (e.g., around 50 seconds) after which low-reliability automated systems have significantly smaller reliance over time than high-reliability systems; before this moment, low-reliability systems have a slightly higher reliance over time than high-reliability systems.

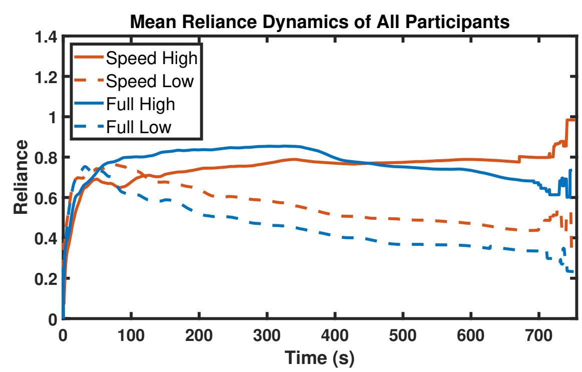


Figure 2. The impact of varied automation conditions on the mean reliance dynamics R^{ac} of all operators.

Reliance Dynamics of Different Performance Groups Across Varied Automation Conditions

We provide a summary of the mean reliance values at the end of a trial, across varied automation conditions and performance groups, in Table 1. According to this table, the order of the automation conditions, from higher to lower reliance, is SH (0.77) > FH (0.68) > SL (0.43) > FL (0.32). This result is in line with that provided in Gremillion et al. (2016). In addition to this, Table 1 shows that a group with a different performance level has a different reliance across those four automation conditions. For example, the Low-Performing Group has the lowest reliance among all groups at Speed Automation (more workload) while this group has a higher reliance than most groups at

Full Automation (less workload). The opposite behavior occurs in the High-Performing Group.

We expand the concept of reliance to our defined reliance dynamics for groups of operators with different performance levels across the four automation conditions. Figure 3 suggests that operators' can have different decision histories, depending on their ability, that leads to different reliance dynamics at varied automation conditions.

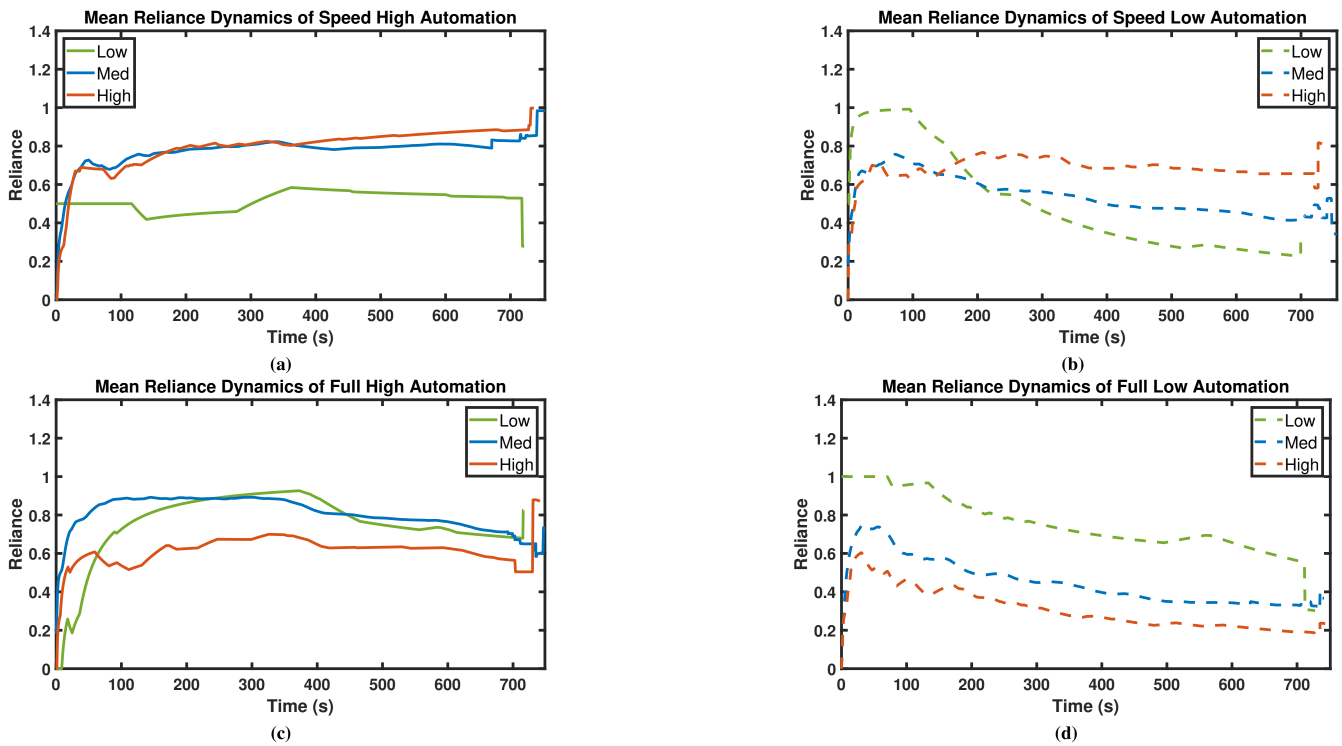


Figure 3. Impact of varied automation conditions on the mean reliance dynamics R^{ac} of different performance groups.

The results indicate that (1) in the SH condition, the average reliance dynamics of High and Medium-Performing groups are similar and higher than the average reliance dynamics of the Low-Performing Group; (2) in the FH condition, the average reliance dynamics of Low and Medium-Performing groups are similar and higher than the average reliance dynamics of the High-Performing Group; (3) in the SL condition, there exists a critical time around 180 seconds, before which the average reliance dynamics are in the order of Low $>$ Medium $>$ High; the opposite ordering happens after 200 seconds; and (4) in the FL condition, the average reliance dynamics are in the order of Low $>$ Medium $>$ High.

Table 1. Mean reliance at the end of a trial by performance groups and four automation conditions. The rows of the table specify the performance group (All, Low, Medium, and High-Performing, respectively), and the columns indicate the automation condition (SL SH, FL, and FH, respectively).

Performance Group	Speed High	Speed Low	Full High	Full Low
All	0.77 ± 0.21	0.43 ± 0.24	0.68 ± 0.20	0.32 ± 0.16
Low	0.53 ± 0.35	0.22 ± 0.12	0.68 ± 0.21	0.55 ± 0.35
Medium	0.79 ± 0.19	0.41 ± 0.22	0.71 ± 0.15	0.31 ± 0.10
High	0.89 ± 0.10	0.65 ± 0.27	0.56 ± 0.39	0.19 ± 0.06

Additionally, (6) the average reliance dynamics of the Medium-Performing Group is similar to that of all operators; this is due to the large number of operators that have been classified in this group; (7) for the Low-Performing Group, after 300 seconds, the order of the average reliance dynamics is FH $>$ FL $>$ SH $>$ SL; and (8) for the High-Performing Group, after about 50 seconds, the order is SH $>$ SL $>$ FH $>$ FL.

The average reliance dynamics orders of both performance groups and automation conditions indicate that the Low-Performing Group relies more on the Full type of automation (less workload) over time, whereas in contrast, the High-Performing Group relies more on the Speed type of automation (more workload).

DISCUSSION AND FUTURE WORK

In human-automation interactions, statistics can provide differential information about operators' reliance on different types and levels of automation in aggregate. However, a dynamical analysis approach can show more sensitive changes in each condition over time to discover more subtle influencing factors. We defined the dynamic process of reliance as the fraction of time a human operator has used automation at a given time. The present study focused on analyzing the impact of factors such as the type of automation (Speed, Full) and its reliability level (Low, High) on the dynamics of reliance on automation for different performance groups. In addition, we studied the impact of reliance choices on the operators' performance across different automation conditions.

The results of analyzing the mean reliance dynamics of all operators across different automation conditions showed that operators tend to rely more on the automation systems with high reliability than on those with low reliability. These findings are consistent with the results addressed by Rodriguez et al. (2021). In addition, we observe that there is a delay for this distinction between operators' reliance in high and low-reliability systems (see Figure 2). In fact, low-reliability systems have a slightly higher reliance over time than high-reliability systems during the first 50 seconds, and it is after this point that the mentioned distinction becomes evident. This behavior may be due to the time that the operators need to notice the automation's performance. So, on average, the operators of the experiment showed an intention of using the automated systems, even without knowing how well these systems were going to perform. They seemed to notice the performance level of the automated systems within the first 50 seconds. That is the point where they started taking over control to improve their overall driving performance. Another interesting finding is that at the end of the trial, and during most of it, the Speed High condition had the highest reliance values, while the Full Low condition obtained the lowest values in such measures. These results indicate that operators expect that partial (Speed) automation would cause fewer penalties than full automation. This explains why, even in low-reliability conditions, people relied more on the Speed type of automation than they did on the Full type.

The analysis by performance groups showed that usually, the High-Performing group has the highest reliance values over time in the Speed type of automation, followed by the Medium, and Low-Performing groups, respectively. The opposite happens for the Full type of automation, where the Low-Performing Group has the highest reliance values over time, followed by the Medium, and High-Performing groups, respectively. These findings indicate that the High-Performing Group finds more productive, performance-wise, to rely more on automated systems that have fewer degrees of freedom because they have a lower chance of making mistakes. This reasoning may have led this group to obtain high final scores. On the other hand, the Low-Performing Group went against this reasoning, and therefore, their score was heavily affected. A similar behavior is observed in the average number of decision changes each group has and the average duration of those decisions. The Low-Performing Group presents fewest decision changes and the largest duration of those decisions in contrast with the Medium and High-Performing groups. This indicates that the Low-Performing Group may have fallen into inappropriate reliance because they kept using the automation for longer periods even though it was not performing well.

Future research includes the spatial dynamical analysis of operators' reliance on automation, the analysis of different perturbations' influence on reliance, and the development of mathematical and machine learning models to determine the dynamics of trust and decision-making for all operators' data.

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

Chiou, E. K., & Lee, J. D. (2021). Trusting Automation: Designing for Responsivity and Resilience. *Human Factors*, (April). https://doi.org/10.1177/00187208211009995

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. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of augmented cognition: Neuroergonomics and operational neuroscience* (pp. 157–167). Springer International Publishing.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Ezer, N., Fisk, A. D., & Rogers, W. A. (2007). Reliance on automation as a function of expectation of reliability, cost of verification, and age. *Proceedings of the Human Factors and Ergonomics Society*, *1*, 6–10. https://doi.org/10.1177/154193120705100102

Gao, J., & Lee, J. D. (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*, 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

Guznov, S., Lyons, J., Nelson, A., & Woolley, M. (2016). The effects of automation error types on operators' trust and reliance. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9740*(1), 116–124. https://doi.org/10.1007/978-3-319-39907-2_11

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

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. Ahram, W. Karwowski, A. Vergnano, F. Leali, & R. Taiar (Eds.), *Intelligent human systems integration 2020* (pp. 999–1006). Springer International Publishing.

Parasuraman, R. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*, 230–253.

Rodriguez, L., Bustamante, C., 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

Ross, J. M., Szalma, J. L., Hancock, P. A., Barnett, J. S., & Taylor, G. (2008). The effect of automation reliability on user automation trust and reliance in a search-and-rescue scenario. *Proceedings of the Human Factors and Ergonomics Society*, *2*, 1340–1344. https://doi.org/10.1177/154193120805201908

Shladover, S. E. (2021). 'self-driving' cars begin to emerge from a cloud of hype. https://www.scientificamerican.com/article/self-driving-cars-begin-to-emerge-from-a-cloud-of-hype/

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors*, *51*(3), 281–291. https://doi.org/10.1177/0018720809338842