2020

The Rise, Fall, and Repair of Trust for Automated Driving Systems

Scott Mishler
Old Dominion University, smish001@odu.edu

Jing Chen
Old Dominion University, j1chen@odu.edu

Original Publication Citation

Mishler, S., & Jing, C. (2020). The rise, fall, and repair of trust for automated driving systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 2006. https://doi.org/10.1177/1071181320641485

The Rise, Fall, and Repair of Trust for Automated Driving Systems

Scott Mishler, Jing Chen
Old Dominion University

The purpose of this study was to investigate how human driver's trust in the automated driving system is built over time and affected by automation failure. The study expanded trust development over time by measuring trust after a practice demonstration of the system capabilities and after each of seven unique, sequential drives. The automation performed perfectly on six of the seven drives but made one of three different responses to a critical hazard event in the fourth drive. Depending on the error-type condition, the automation either perfectly avoided the hazard (no error), issued a takeover request (TOR), or failed to notice the hazard (failure). In contrast to the typically used pre/post trust-difference assessment that does not show a trajectory of growth or decline patterns, the current design allowed for evaluation of the growth, decline, and repair of trust.

In the practice, a demonstration of the automated driving system, pre-drive, was done to allow us to assess initial learned trust. Drive number (pre-drive and drives 1-7) was the within-subjects independent variable and automation error type (no error, TOR, failure) was the between-subjects independent variable. Subjective trust was the dependent variable. An overall increase in trust was shown during the pre-drive and drives 1-3, which demonstrates that trust increases as participants gained more experience with the system. Because the system was perfect in Drives 1-3, the increase over time indicates proper trust calibration (Hancock et al., 2011; Lee & See, 2004; Parasuraman & Riley, 1997). This initial development of trust is a promising display of how trust can grow in an initial experience with a reliable automated driving system.

For drive 4, trust was significantly higher in the no-error condition compared to the TOR and failure conditions, with no significant difference between the latter two. The TOR was expected to decrease trust less than the failure because it was a warning, not a complete failure. Yet, this pattern was only shown numerically with no statistical significance. During the three drives after the critical drive 4, trust was slowly repaired for the TOR and failure conditions. However, the increase did not continue upward, but remained at the same level during the last two drives. This result indicates that although trust rebounded after a critical event, it still did not reach previous trust levels where no problem had occurred. TORs and automation failures can have a lingering negative effect on trust, potentially harming later human-automation collaboration. Participants likely did not know the exact cause of the problem and could have worried that the same thing could happen again, with no way to predict it. Because this automation did not explain why the TOR was issued, it lacked transparency. Therefore, participants likely assumed the automation had just failed. Providing transparency information and explaining errors can help trust resilience (Dzindolet, et al., 2003, Hoff & Bashir, 2015).

Results of the current study imply that designers should pay careful attention to the amount of time drivers have spent with automation because trust could take longer to develop than previously expected. Designers should also consider trust impairment caused by automation errors and potential strategies to repair trust.

References

Chen, J., Mishler, S., Hu, B., Li, N., & Proctor, R. W. (2018). The description-experience gap in the effect of warning reliability on user trust and performance in a phishing-detection context. *International Journal of Human-Computer Studies*, *119*, 35-47.

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

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*, 53–71.

Röttger, S., Bali, K., & Manzey, D. (2009). Impact of automated decision aids on performance, operator behaviour and workload in a simulated supervisory control task. *Ergonomics*, *52*, 512-523.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science*, *2*, 352-367.