Trust in Automated Vehicles: The Roles of Degree and Transparency

The National Highway Traffic Safety Administration has stated that fully autonomous vehicles will become a reality in the not-so-far future (United States Department of Transportation, 2020). While most vehicles today cannot drive themselves, a wide variety of automated driving technology has become available to the average consumer. These systems range from lane-change assist to the autopilot features found in Tesla. With the rise of assistive driver technology in cars, it is important to understand how drivers will interact with them. One factor that has a profound influence on the ways people interact with automated systems is trust (Lee & See, 2004). A user's trust in an automated system will dictate how motivated they are to follow the advice of a system or use it to its fullest potential. If users do not trust a system, they may be more likely to override the safety features present in the system, and if users trust a system too much, they may fail to use manual control when it is necessary (Banks & Stanton, 2016).

When assessing trust in automation, it is important to distinguish the degree of automation you are referring to. Typically, lower levels of automation require more input from the human user, while higher levels of automation require more input from the automation (Endsley, 2017). Parasuraman et al. (2000) classified automated systems into 4 different categories: Information acquisition, information analysis, decision and action selection, and action implementation. The latter three types are of particular interest in the driving domain, as they encompass the most recent automated technology that is available on the market today and the automation that will be available in the near future.

Another known characteristic of automation that has influence on users' trust is the level of transparency the system provides. Transparency can be thought of as the amount of

information a system provides about its decision-making process to the user of the system (Bhaskara et al., 2020). In their review of automation transparency, Bhaskara and colleagues state that increasing transparency should allow the operator to develop a more accurate mental model of the system and its behavior. A study by Mercado et al. (2016) investigated how different levels of transparency affected user trust and compliance with an autopilot system on an Unmanned Aerial System (UAS). They found that participants' subjective ratings of trust in the system increased as the amount of transparency provided by the automation increased. However, according to Cognitive Load Theory, too much extraneous information may confuse the user and lead to more errors (Kalyuga et al., 1999). Although transparency appears to have a generally positive impact on perceived usability and trust, some research suggests that the appropriate degree of information may be context specific (Bhaskara et al.); that is, greater transparency may be valuable in some situations and detrimental in others.

To understand how users are likely to perceive an automated system, then, it is important to examine the effects of transparency in relation to the degree of automation. This allows us to more precisely identify the circumstances in which users are likely to trust an automation and circumstances in which trust may be limited. In the present study, we investigate the difference in drivers' subjective ratings of trust in an automated vehicle system across different degrees of automation and transparency. Specifically, we explored trust in reaction to three different levels of driving assistance automation: driver controlled turn-by-turn navigation, supervisory control autopilot, and totally autonomous autopilot that gives no control to the driver. As we were interested in how ratings of trust vary across differences in transparency, we also varied whether each automation was transparent or non-transparent in explaining decision processes to the

driver. Finally, as past research has indicated an association between extraversion and trust in automation (Merritt & Ilgen, 2008), we assessed participants' extraversion.

Method

Participants

We recruited adult volunteers through paper flyers and online advertisements in the communities surrounding two campuses of a large state university. One campus is located in a mid-sized traditional college town and the other is located in the downtown area of a large city. To recruit as broad a sample as possible, we posted advertisements in local restaurants, gyms, and community centers and purchased online ads on the pages of two local news sites. When participants responded to the advertisement, we verified that they had a current driver's license and then invited them to a single two-hour research session in the driving simulator laboratory on the campus most convenient for them. There were no other inclusion or exclusion criteria. At the beginning of this session, we explained the study tasks, risks, and compensation and gave them the opportunity to ask questions. Participants who gave informed consent to participate then began the study. At the conclusion of the study, we paid each participant $50 and reimbursed their expenses for parking or public transportation..

Materials

Driving Simulator

Participants completed 20 trials, each lasting 5 minutes, in a realistic driving simulator. The simulator technology was identical in both labs. The virtual routes in the trials varied, including a busy urban environment, an interstate highway, and a winding mountain road. All participants completed the same 20 routes, but we randomized the order of the routes across participants to prevent ordering effects. The simulator mimicked the driver's seat of a passenger

car and showed realistic video of the route on a windshield-sized screen. We varied the degree of automation across three conditions. In the driver control condition, the simulator provided turn-by-turn GPS directions, but the participant controlled the vehicle completely. In the supervisory control condition, the simulator drove via an autopilot feature that participants could override when they felt necessary, giving them control over the vehicle when they chose. The final level of automation was a fully autonomous autopilot feature that did not allow the driver to make changes to the vehicle's course or overtake control in any way.

Transparency

For half the participants in each automation condition, we made the logic behind the system's decisions transparent to the participant. Whenever the vehicle needed to execute a maneuver such as turning, stopping, or speeding up, the system would give an auditory explanation such as "we're on the fastest route," "this will allow you to avoid a construction zone," or "stop sign ahead". Participants in the nontransparent conditions did not hear these audio messages.

Measures

Trust in Automation

To assess the participants' trust in the automated system, each participant completed the Merritt Scale (Merritt, 2011) at the conclusion of the driving simulation. This is a six-item questionnaire designed to assess a user's attitudes toward an automated system. Participants respond to their level of agreement with each statement on a 5-point Likert Scale ranging from 1- "strongly disagree" to 5- "strongly agree". One example of an item on the Merritt Scale is: "I can depend on the system". We summed these scores to create an overall index of trust ranging from 6 to 30.

Extraversion

We assessed participants' extraversion using Goldberg's (1992) measure from the International Personality Item Pool. Each item consists of a statement (ex. "I… am the life of the party") and participants rate the degree to which they agree that each item describes themselves on a 7-point scale, with 1 being "strongly disagree" and 7 being "strongly agree."

Procedure

After participants consented to participate in the study, we asked them to complete the measure of openness. We then randomly assigned each participant to a level of automation (decision aiding turn-by-turn, supervisory control, and fully autonomous) and to a transparency condition (transparent automation, or nontransparent automation). All participants viewed a brief instructional video that explained the driving simulation task and the automated system they would be interacting with. Each participant then completed 20 driving trials, each of a different simulated course, lasting 5 minutes each. After completing the driving trials, participants completed the Merritt Scale. We provided a short verbal debriefing about the goals of the experiment, then provided participants with their compensation and thanked them for their input.

References

Banks, V. A., & Stanton, N. A. (2016). Keep the driver in control: Automating automobiles of the future. *Applied Ergonomics, 53* (B)*,* 389-395. https://doi.org/10.1016/j.apergo.2015.06.020

Bhaskara, A., Skinner, M., & Loft, S. (2020). Agent transparency: a review of current theory and evidence. *IEEE Transactions on Human-Machine Systems*, *50*(3), 215-224. https://doi.org/10.1109/THMS.2020.2965529

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human automation research. *Human Factors, 59*(1), 5–27. https://doi.org/10.1177/0018720816681350

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351-371. https://doi.org/10.1002/(SICI)1099-0720(199908)13:4%3C351::AID-ACP589%3E3.0.CO;2-6

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for multi-UXV management. *Human Factors*, *58*(3), 401-415. https://doi.org/10.1177/0018720815621206

Merritt, S. M. (2011). Affective processes in human–automation interactions. Human Factors, 53(4), 356-370. https://doi.org/10.1177/0018720811411912

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors, 50*(2), 194-210. https://doi.org/10.1518/001872008x288574

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, 30*(3), pp. 286–297. https://doi.org/10.1109/3468.844354

U.S. Department of Transportation. (2020). *Automated vehicles for safety*. Retrieved from https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety