Decision making in human-autonomous vehicle interaction

Anirudh Unni[1], Alexander Trende[2], Claire Pauley[1], Bianca Biebl[3], Severin Kacianka[4], Andreas Lüdtke[2], Klaus Bengler[3], Alexander Pretschner[4], Jochem W. Rieger[1]

[1] *Department of Psychology, Oldenburg University, Oldenburg, Germany*

[2] *OFFIS Institute for Information Technology, Oldenburg, Germany*

[3] *Chair of Ergonomics, Technical University of Munich, Garching, Germany*

[4] *Chair of Software and Systems Engineering, Technical University of Munich, Garching, Germany*

Introduction and Goal

A majority of vehicle accidents is caused by human errors (Singh, 2015). The main promise of autonomous driving is that autonomous vehicles (AV) will reduce traffic accidents caused by human errors and will eventually be safer than human driven vehicles (HV). However, until a time comes when only fully autonomous vehicles travel on roads, the interaction between HV and AV remains extremely important. There is evidence that humans apply different moral concepts for interactions with cyber-physical systems which are much more outcome-oriented as compared to norm-oriented interactions with other humans (Malle et al., 2015). This implies that humans will interact differently with technical systems, like AV than with HV. The aim of this study was to investigate if there is a difference between the valuation of actions when an interaction involves AV as compared to similar interactions with other HV in time-critical merging situations and whether these potential differences in human-human and human-autonomous vehicle interactions can be characterized from behavior and neurophysiological whole-head fNIRS brain activation measurements.

Methods and Data Analyses

In order to investigate differences in human-human and human-autonomous vehicle interactions, we performed a within-subject driving simulator study where we presented 12 participants (21-29 years, mean ± SD = 23.8 ± 2.61) with a cover story that they were driving in a time-critical situation through urban traffic in the presence of other HV and AV. Additionally, we mentioned that the AV were programmed defensively and drove conservatively to avoid collisions and had faster braking reaction times than HV as this is the expected programming of AV (Li & Sun, 2018; Zhan et al., 2016). However, the AV and HV simulations actually presented the same driving behavior. Participants were told that they would receive a monetary reward if they managed to finish the driving block within a given time limit while avoiding risky driving maneuvers. On their way, participants were repeatedly confronted with a left-lane merging situation at unsignalized intersections (10 intersections per driving block) where they had to decide to merge in front of HV or AV. Participants drove 10 such driving blocks (100 intersections in total). The given time limit and the traffic density within the intersection were designed such that participants had to merge quickly to finish the block in time to receive the bonus, thus forcing a decision to merge in front of an oncoming AV or HV. This period where the participant stops at the intersection and later initiates the left-lane merging maneuver was defined as the decision-making phase. We recorded whole-head fNIRS data throughout the driving task.

To explore if there were any behavioral differences in merging between AV and HV, we fitted a logistic model to the gap acceptance probability of the participants for AV and HV separately and derived a model threshold indicating a 50% gap acceptance for AV and HV.

After pre-processing the fNIRS data and reducing physiological artefacts, we performed two types of analyses. The first was a multivariate decoding modeling framework where we decoded whether the participant currently decided to merge in front of a HV or AV from the fNIRS HbR activity using a binary multivariate logistic ridge regression model within a 5-fold cross-validation. In the second analysis, we characterized the contribution of the brain activation features to such a decoding model that predicted human-human (merging in front of HV) or human-autonomous (merging in front of AV) interactions on a group-level by reporting the effect sizes for each fNIRS channel. For this, we first performed a channel-wise paired t-test from the preprocessed fNIRS data for the two conditions AV and HV. We calculated Cohen's d for each channel from t-values to indicate the effect sizes in sensor space. The Cohen's d brain maps are reported for the group-level analyses.

Results

The models' threshold values indicating the 50% gap acceptance for AV and HV were m_{AV} = 3.09s (2.96s – 3.20s) and m_{HV} = 3.08s (2.75s – 3.30s) respectively. The threshold values and their corresponding confidence intervals show an overlap, suggesting that these distributions do not differ significantly. However, the neurophysiological results indicate that our approach of using whole-head fNIRS in combination with a cross-validated multivariate logistic ridge regression could be useful to decode the interaction partner at a crossing, i.e., whether the participants decided to merge in front of HV or AV. The decoding accuracy was statistically significant as compared to the 95% CI for the empirical chance level in all participants. This approach allowed us to exploit the spatial specificity of whole-head fNIRS, that enabled us to predict the interaction partner at the crossing with an average accuracy of 67.2% (SD = 5%) across all participants and up to a maximum of 75.9% on a single-subject level.

The Cohen's d brain maps reporting the contribution of the brain activation features to the decoding model are shown in Figure 1. The results from this group-level analyses showed the largest effect sizes of brain activation differences between the merging phases for AV and HV (AV-HV) in the prefrontal cortex, especially in the left and right dorsolateral areas and left ventrolateral prefrontal areas (Cohen's d $\sim$ 0.9 to 1.2). Additionally, the ventromedial prefrontal areas and the superior frontal gyrus indicated increased activation differences for the decision-making phase during merging in front of AV as compared to HV (Cohen's d $\sim$0.9). These neural correlates from the prefrontal areas have been previously represented in values of actions taken during decision-making (Hollmann et al., 2011; Gläscher et al., 2009; Sanfey, 2007).

Discussion and Conclusion

Our results provide evidence that humans show differences in the decisions they make in time-critical situations depending on whether they interact with AV or HV and this is expressed in fNIRS brain activation differences even though there are no observable behavioral differences. Overall, our results demonstrate a consistent difference in activation at the brain-level and these activation differences occur in brain areas that have been previously related to valuation of actions during decision-making.

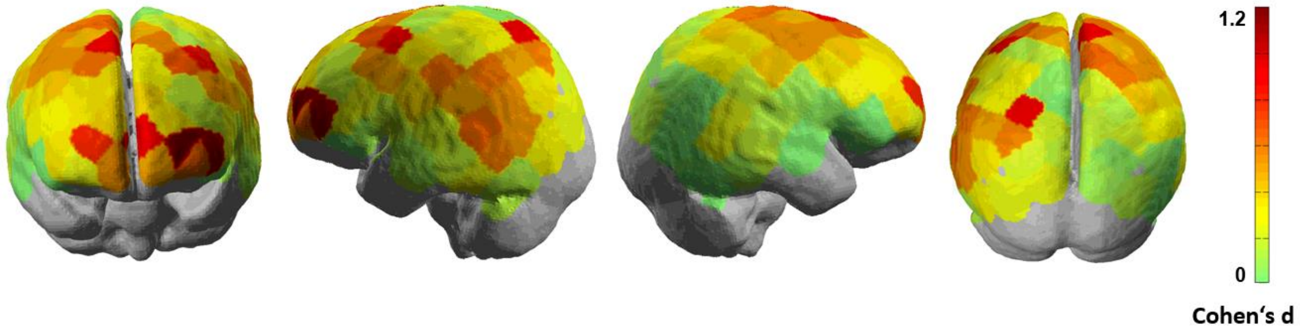

Figure 1: Cohen's d brain maps representing effect sizes computed from channel-wise weighted averaged t-statistics for the group level analyses. Moderate to high Cohen's d values (0.8-1.2) show medium to large effect sizes indicating increased activation differences for the decision-making phase during merging in front of AV as compared to HV.

References

Hollmann, M., Rieger, J. W., Baecke, S., Lützkendorf, R., Müller, C., Adolf, D., & Bernarding, J. (2011). Predicting Decisions in Human Social Interactions Using Real-Time fMRI and Pattern Classification. *PLoS ONE*, *6*(10), e25304. https://doi.org/10.1371/journal.pone.0025304

J, Gläscher., AN, Hampton., & JP, O'Doherty. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex*, *19*(2), 483–495. https://doi.org/10.1093/CERCOR/BHN098

Li, X., & Sun, J. Q. (2018). Defensive driving strategy and control for autonomous ground vehicle in mixed traffic. *Studies in Computational Intelligence*, *731*, 3–44. https://doi.org/10.1007/978-3-319-64063-1_1

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015). Sacrifice One for the Good of Many?: People Apply Different Moral Norms to Human and Robot Agents. *ACM/IEEE International Conference on Human-Robot Interaction*, *2015-March*, 117–124. https://doi.org/10.1145/2696454.2696458

Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. In *Science* (Vol. 318, Issue 5850, pp. 598–602). American Association for the Advancement of Science. https://doi.org/10.1126/science.1142996

Singh, S. (2015). Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. *Traffic Safety Facts - Crash Stats*.

Zhan, W., Liu, C., Chan, C. Y., & Tomizuka, M. (2016). A non-conservatively defensive strategy for urban autonomous driving. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 459–464. https://doi.org/10.1109/ITSC.2016.7795595