Evaluation of automated vehicle behavior in intersection scenarios

Thomas Streubel, Pierre De Beaucorps, Fawzi Nashashibi

Evaluation of automated vehicle behavior in intersection scenarios

Thomas Streubel, Pierre de Beaucorps and Fawzi Nashashibi

Robotics and Intelligent Transportation Systems (RITS), INRIA Paris, 2 rue Simon Iff, 75012 Paris, France

thomas.streubel@gmail.com, pierre.de-beaucorps@inria.fr, fawzi.nashashibi@inria.fr

Abstract

The development of automated vehicles (AVs) is ongoing and soon the first versions will enter the up to now exclusively human controlled traffic environment. In previous research, we developed an approach for an automated vehicle control in intersection scenarios. The focus was on the decision-making process to either enter the intersection before another car or wait to let the other car pass. While objective risk features were used to evaluate the performance, the interaction with real human drivers remains uncertain. Therefore, we propose a method to evaluate the automated vehicle behavior from an outside perspective by introducing a human driver and our automated vehicle in the same simulation environment. We conducted a study to examine the interaction and evaluate the automated vehicle behavior from another driver's perspective. The focus of the study was on the naturalness and risk of the behavior and how the subjective risk evaluation is linked to our main objective risk feature.

For comparing the results, we used two different setups for our automated vehicle (aggressive vs. passive) and varied the starting position to create either of the two situations (AV first vs. driver first) as mentioned above. The results show a high overall naturalness score. The risk score was related to the outcome of the situation and the AV setup. A weak correlation between the subjective risk assessment and the objective risk feature indicated the necessity to evaluate automated vehicle control from an outside perspective to achieve a better assimilation in human traffic environments. Since the identification of automated vehicles can influence the driver behavior, the integration of a human-like behavior for AVs seems highly desirable.

Keywords

Multi-agent simulation; subjective risk assessment; automated vehicle behavior; automated vehicle evaluation

1. Introduction

On the verge of automated vehicles (AVs) entering the existing traffic environment new challenges arise. This concerns especially the interaction with human driven vehicles which is subject of present research. Here, the focus is mainly on the reaction of the automated vehicle towards other human controlled cars in real traffic environments and particularly the performance within certain comfort limits to satisfy passenger's expectations inside the AV. The performance from an outside perspective especially its ability to "blend in" is less investigated. However, the development of automated vehicle control requires intense testing and validation when it comes to interaction with other vehicles. This is commonly conducted against predefined safety and comfort metrics but also requires subjective testing when integrated into a human domain such as the traffic environment. The latter point is particularly relevant when the presence of an automated vehicle is influencing the human drivers' behavior [1].

In longitudinal traffic, this influence was shown in [2]. A low time headway (THW) in a platoon of automated vehicles led to a reduction of the THW by the driver even below a safety threshold of 1.0s. This behavior adaptation is safety critical and requires more awareness especially since modelling driving behavior is only based on the current state of human controlled traffic environments.

We propose a method to evaluate the automated vehicle behavior in intersection scenarios. Intersections are challenging elements of the traffic network with a high level of interaction between traffic participants leading to high numbers of accidents. Due to this danger, intersection scenarios with AVs are best investigated in simulation. Therefore, a simulation environment was created where human drivers can interact with others and also with automated vehicle. In previous research, we developed a setup for an automated control focusing on the decision making process at intersections [3]. The situation when two vehicles approach an intersection can dissolve into two possible outcomes (besides a collision). We use the post-encroachment time (PET) as a risk feature for the decision process. This time refers to a conflict zone that both approaching vehicles will pass when crossing the intersection. The PET is mentioned in [4] as a parameter related to risk for left turning scenarios. It is defined as the time between one vehicle leaving the conflict zone and the other one entering this zone. The study examined the driving behavior in a real driving scenario under two conditions of the driver (comfort vs. hurried) to explore comfort limits of the PET. The results showed mean values between 1.5s (hurried condition) and 2.2s.

In this paper, we investigate the interaction between one human driver and an automated vehicle at an X-shaped intersection. We conducted a simulator study to evaluate the performance and behavior of our automated vehicle subjectively and compare this with our objective risk evaluation. The naturalness and the risk assessment of the human drivers towards the AV were focused on. As objective risk feature, we use the PET.

2. Methodology

Simulation environment

The simulation environment used in this study is self-designed in JavaScript as a multi-agent browser application with client-server architecture. This way, several drivers can interact in the same simulation environment as well as with our automated vehicle. The simulator tool is easily accessible through a browser without any installation. A 3D view and a bird-view are shown in Figure 1. In the drivers perspective the assigned path is highlighted. To reduce the complexity of the situation, the human controlled car was always going straight at the intersection.

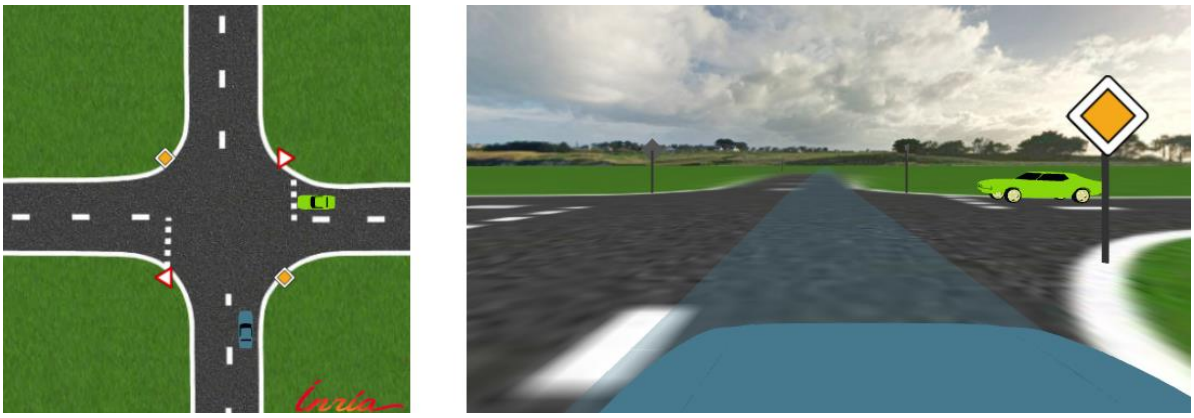

Figure 1: bird-view (left) and in-vehicle perspective (right) of the simulation tool

The simulation is based on the open source bullet physics engine [5]. The vehicle model is an adaptation of the Raycast vehicle [6] with a suspension system for each wheel and a body mass as chassis above. This realizes pitch and bounce in the simulation giving a realistic impression of rapid changes in speed to the human driver. The computation of the physical motion of the vehicles is centralized on a server while the client hardware is only responsible for the visual representation. This saves computation resources and enables the usage of the simulation with multiple clients.

For the control of the vehicle, there are a steering wheel and pedals for realistic inputs. However, while the steering wheel was set up in front of the participants to create a car like atmosphere, it remained unused because of the predefined straight path. The lateral control was automated since the usage of steering within the simulation requires some training and fine-tuning. The pedals were used to control the acceleration and brake. Shifting gears was unnecessary since the vehicles are considered automatic. For the visual output, we used three monitors arranged slightly concave. This realizes a viewing angle of about 114 degrees. Rearview mirrors are not implemented.

Participants

Overall 16 drivers participated in the simulator study. Three of them were female. The average age was 30.4 years (*sd*= 9.1 years). Everyone was in possession of a valid driver's license in average for 10.1 years (*sd*= 7.3 years). The mean annual travel distance was around 6,000 km with about half of it in urban environments.

Scenarios

We created a standard X-shaped 4-way crossroad with one lane entering and exiting in each direction. The test drivers were passing the intersection going straight. Hence, there are six scenarios how an automated vehicle could interact with the drivers by interfering with their path (cf. Figure 2). The conflict can arise either by crossing paths or by merging into the same direction which can result in different incidents. The scenarios are distinguished by the maneuver of the AV combined with the potential conflict: left turn across path (LTAP), left turn same paths (LTSP), right turn same paths (RTSP) and straight across path (SAP). The approaching direction relative to the human driver is the second criteria to describe a scenario: left direction (LD), right direction (RD) and opposite direction (OD). All combinations with interaction result in the six scenarios: two for left turn across path (LTAP-RD and LTAP-OD), one left turn same paths (LTSP-LD), two for straight across path (SAP-RD and SAP-LD) and one right turn same paths (RTSP-RD) scenario (see Figure 2).

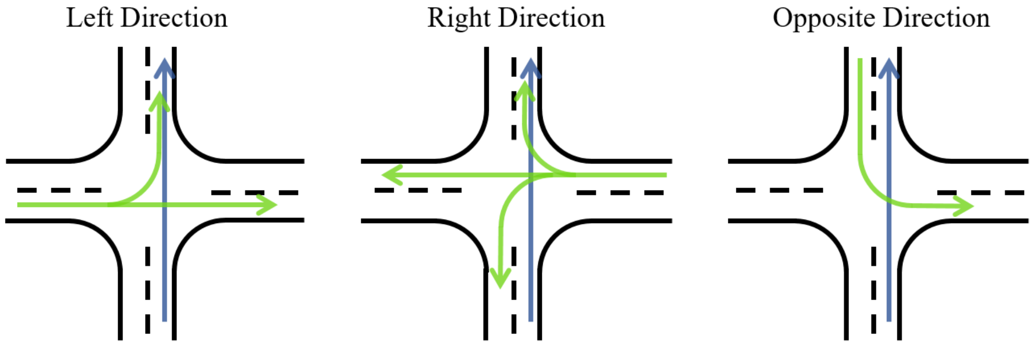

Figure 2: Scenarios grouped by direction of the AV (green) relative to the human driver (blue)

The traffic signs were placed so the drivers had always the priority. Hence, the automated vehicle should yield unless passing before the driver is safe and is not interfering with the trajectory of the human driver. These situations were especially of interest in the study since a car entering in front of the human can be experienced as unnatural or risky especially when it forces a reaction from the driver. To create those situations with interaction the starting position of the AV was varied.

Automated Vehicle Control

A detailed explanation of the automated vehicle control and decision algorithm is given in [3]. Here, we give a brief overview. The decision algorithm includes a simple prediction of the other vehicle. Further, we extracted several speed profiles from real drivers driving in our simulation. These speed profiles are clustered and serve as proposals for the decision algorithm. The predicted trajectory of the AV is calculated by utilizing these reference speed profiles as target speed taking into account the actual position relative to the intersection and the present velocity. Here, the path is preset since the scenarios are predefined in the study design (see Scenarios). The retrieved trajectory options are now compared to the extrapolation of the other vehicle (constant speed assumption). The prediction horizon is set to about 3 seconds. When both vehicles approach the intersection, the predicted future vehicle positions intersect and a PET value can be estimated. This is the main risk feature in the algorithm.

For the scenarios where both vehicles drive in the same direction after passing the intersection, the definition of the conflict zone is extended since the paths are merging. Here, the conflict zone to determine the PET is starting when both vehicle positions meet for the first time but then is limited to a car length. Without this restriction, the conflict zone would extend over the entire further lane since both cars drive on the same leg of the intersection.

A positive PET is referring to the AV passing before the other vehicle. If the estimated PET value drops below a certain threshold, the associated trajectory is discarded. The decision algorithm is eventually choosing the trajectory with an estimated PET value higher than the threshold and with the highest overall velocity to optimize the travel time. Every time step (about 100ms) all trajectories are tested this way. For cases where all speed profiles are risky, there is a stopping trajectory as fall back option. As soon as the vehicle enters the conflict zone, it passes the intersection quickly and switches into a car following (ACC) mode. This prevents dangerous situations by stopping in the intersection area or following the driver in a too close distance.

Procedure

At the beginning of the study the participants received the instructions and filled out a short demographic survey. Afterwards, they were able to get used to the vehicle control and the simulation environment in a small test track. When the driver felt comfortable with the pedals, the study started. The intersections were presented to the driver consecutively in a randomized order to avoid sequence effects. As mentioned before, the interaction between AV and driver was desirable. However, since the approaching behavior of the drivers is dependent on individual preferences, a synchronization to standardize the scenarios is nearly impossible. The AV was early visible in our open field design, so late manipulation of the AV could affect the assessment of the driver and also would be in contradiction to the algorithm for the automated vehicle control that aims to avoid high risk situations.

However, to avoid delays at the start, the AV is triggered by the first input of the driver (usually pushing the gas pedal). To realize different situations with AV passing before the human driver and also yielding, we chose three different starting positions for the AV fine-tuned for every scenario. Finally, the simulation stops shortly after the driver passed the intersection. Since the path was straight, the driver could focus on the longitudinal control with the driving pedals. The path was also visible in the scene for the driver to assure the driving direction (see Figure 1 right side). After each scenario, the drivers were asked about the naturalness of the behavior of the other vehicle, how risky they rank this behavior and if they would change their own behavior retrospectively.

Design

The study was conducted in a repeated-measurement design. All six scenarios were presented to every driver in a randomized order. There were two different automated vehicle setups (aggressive vs. passive) to vary the behavior of the AV. This was realized by changing the PET threshold in the algorithm. The aggressive setup accepts PET values down to 0.5 seconds while the passive one only passes before the human driver if the estimated PET is above 1.5 seconds. This value is chosen in compliance with the results in [4] for the hurried human driver. Previous tests have shown that even higher thresholds lead to a very conservative behavior and almost always result in a yield to oncoming vehicles.

The starting distance of the AV was varied between three points to create interaction in the scenarios. Further, the vehicle dynamics were recorded for every run enabling a replay afterwards to confirm certain actions. For the evaluation, the actual PET was determined for every case. Here, a positive PET indicates that the AV was passing before the driver and negative PET values imply the opposite. Accordingly, low absolute PET values can be seen as more dangerous than higher absolute values. In the short questionnaire after every setup, the naturalness and riskiness of the other vehicle was ranked by the test drivers.

3. Analysis and Results

The driving data along with the answers to the questionnaire were stored during the study by the simulation tool into json-files. For the analysis, this data is converted into Matlab. Overall, the test drivers crossed the intersection 576 times (16 drivers * 2 AV setups * 6 scenarios * 3 variations). The three variations refer to different starting points for the AV to create interaction in the scenarios. In the process, the AV passed 235 times first (40.8%). This situation occurred more often in the aggressive AV setup as shown in Table 1. There were two incidents with the aggressive AV in which the PET values were zero; one was an actual collision where the AV slightly touched the passing human driver and the other was in the scenario RTSP-RD when the AV entered the designed conflict zone promptly behind the human driver.

Table 1: Order of passing according to AV setup

AV setup	AV first	driver first
Aggressive	140	146
Passive	95	193

Risk

For the objective risk evaluation and as risk feature in the automated vehicle, we use the PET value. Positive values represent the AV passing before the human and negative values the opposite. In Figure 3 the distribution of the PET values is shown according to the aggressive and passive AV setup. The bars are semitransparent to display both AV setups in one graph. The gap after zero indicates that almost no situations appeared where the AV entered directly before the human driver. Exceptions are the two incidents described before and one appearance of a PET value slightly above 0.5 seconds in the aggressive AV setup. In the majority of cases, the AV yielded and entered the conflict zone after the drivers.

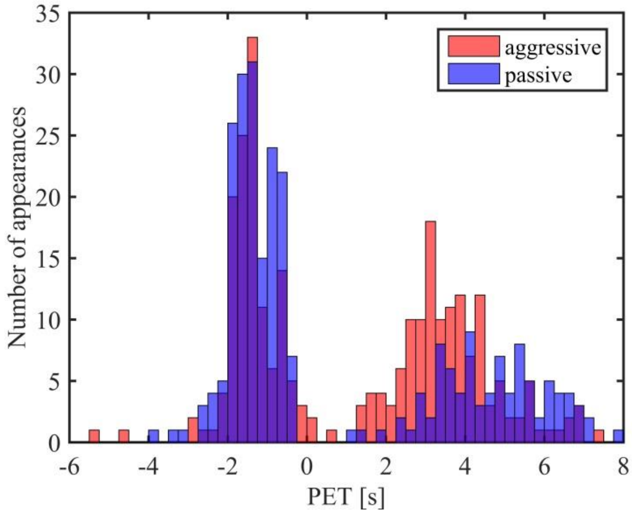

Figure 3: Histogram of the PET values for both AV setups (semitransparent bars)

The small negative PET values indicate that the AV entered the PET zone directly after the driver left which the driver was not able to percept in our setup (missing mirrors in the simulation). So, this should not have influenced the subjective risk assessment but seems not desired.

The subjective risk was assessed by the driver on a ten point scale after every run. Here, the drivers evaluated the riskiness of the behavior of the automated vehicle. The average risk values per scenario and AV setup are shown in Table 2. Below the mean values, there are the confidence intervals on a confidence level of 95%.

Table 2: Mean risk scores per scenario in dependence of AV setup

AV setup		LTAP-RD	LTAP-OD	LTSP-LD	SAP-RD	SAP-LD	RTSP-RD
aggressive	Mean	3.4	4.3	4.4	4.9	4.4	3.1
aggressive	CI	0.66	0.77	0.91	0.95	0.77	0.71
passive	Mean	2.9	2.5	2.9	3.7	3.3	1.9
passive	CI	0.78	0.49	0.75	0.88	0.63	0.38

As expected, the mean risk for the aggressive setup is higher than for the passive setup. A main factor of influence is the change in the situation which was not controlled in the simulation design. So, there is a difference in the rating when the AV enters before the driver compared to when it yields and stops at the intersection. Latter is usually considered less risky and scores lower on the scale accordingly. The low value for RTSP-RD with a passive AV is related to this issue. There are only 14 “AV first” situations (29%) in this scenario. This could explain the low risk rating. Highest rating is SAP-RD with a score of 4.9. Here, there are 36 “AV first” occurrences (75%) which is increasing the risk score noticeably.

The dependency on the situational outcome is further investigated by comparing the risk score according to the situation and the AV setup (see Figure 4). Here, an unbalanced two-way ANOVA was performed showing a main effect in the AV setup ($F(1, 570) = 20.79$, $p < .001$) as well as in the outcome of the situation ($F(1, 570) = 33.79$, $p < .001$). Overall, the risk values remain in a lower part of the scale.

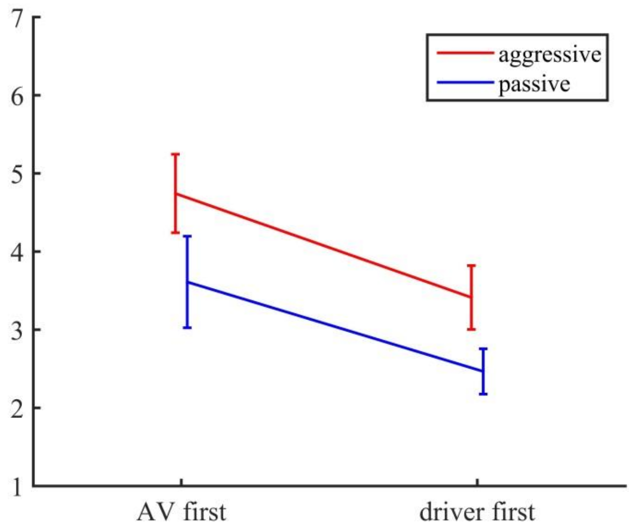

Figure 4: Risk score AV first vs. driver first with both AV setups

For investigating the relationship between our objective risk measure and the risk score, a correlation analysis was performed. We took only the samples with positive PET (AV first), since the driver could not perceive vehicles passing behind the car. There was only a weak correlation (r= -0.42). A closer look into the data of some test drivers revealed that there might have been a misunderstanding of the risk score.

Naturalness

After each passing of the intersection, the driver was asked to rate the naturalness of the behavior of the present vehicle on a 1-10 scale. It was not mentioned that the present vehicle is controlled automatically. The mean score over all crossings was considerably high with 7.3 (sd= 2.56).

Further, the score was evaluated for each of the two AV setups (aggressive vs. passive) and separately for each of the six scenarios. The results are shown in Figure 5. Here, the mean scores and their confidence intervals are displayed. It shows that the scores for the aggressive setup are overall lower than for the passive setup. The LTAP-OD scenario stands out with a very low average of 5.9 (CI= .75). High scores are reached for the passive AV setup in the LTSP-LD scenario (mean= 8.4; CI= .56) and in the RTSP-RD scenario (mean= 8.6; CI= .51).

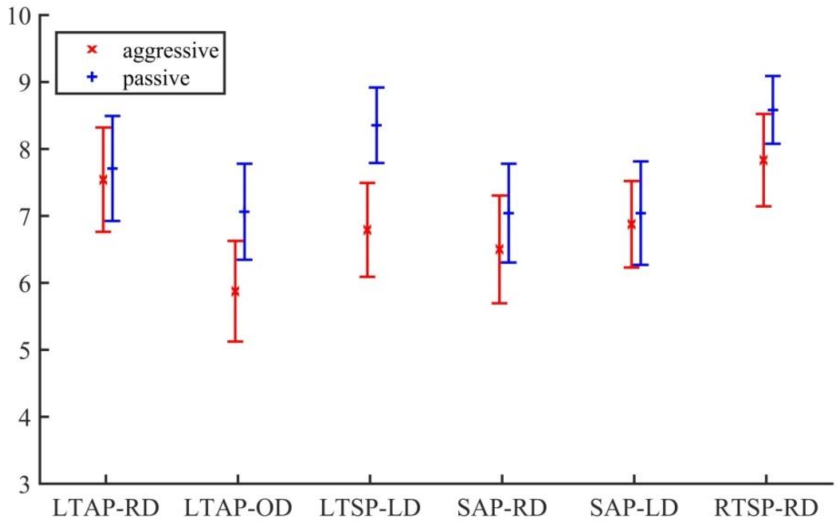

Figure 5: Naturalness score per scenario with both AV setups

The high scores seem to be related to the numbers of the AV passing before the vehicle. The outcome of the situation and both AV setups is compared and shown in Figure 6. An unbalanced two-way ANOVA shows that there is a main effect in the AV setup ($F(1, 570) = 12.9$, $p < .001$). Accordingly, the passive vehicle setup is rated more natural than the aggressive one.

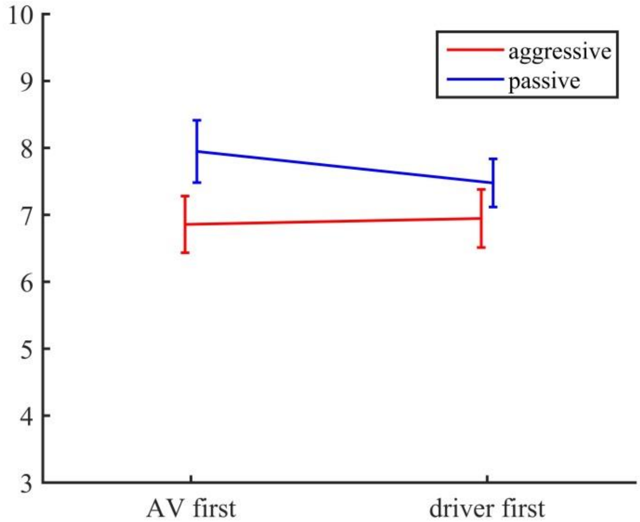

Figure 6: Naturalness score AV first vs. driver first with both AV setup

These results suggest that in the ranking there is a relationship between the risk and the naturalness ranking by the human drivers. This is confirmed by testing the correlation of all sequences between both scores (r= -0.66). However, the objective risk in form of PET shows a weak correlation with the naturalness score for positive PET values (r= 0.24).

4. Discussion and Conclusion

The objective risk evaluation using PET showed two dangerous incidents where the automated vehicle entered too close behind the human driver and even collided once. We identified this weakness of our approach which also leads to the high rate of very low negative PET values in Figure 3. Since the decision-making algorithm is only using the predicted position of the other vehicle to determine the risk of a possible trajectory, the vehicle remains unconsidered as soon as it exits the conflict zone. Basically, a memory attribute is required for the AV to realize that the conflict zone was recently occupied and shall not be entered before a safety margin.

The subjective risk evaluation indicated that both AV setups are distinguishable and the aggressive setup is assessed as more risky in its behavior. We also showed that this is mainly related to the higher numbers of situations where the AV enters the intersection first. The weak correlation between the objective and subjective risk measures shows that the subjective evaluation is reasonable and extends the existing validation. Further, there might be another risk measure that is more related to the human perception which could be investigated in future research.

The overall naturalness score was in the upper part of the scale. In the scenario comparison, the LTAP-OD was rated less natural which is explainable by the missing visualization of blinkers in the simulation. So, the drivers were confused of the vehicle stopped on the opposite site of the intersection or in the other situation turning left and passing in front of them. Apart from that, the high scores for passive AVs in the two scenarios where both vehicles go in the same direction after passing the intersection are interesting. There is most likely a relation to the amounts of situations where the AV yields rather than passes in front.

Showing the correlation between naturalness and the subjective risk score, indicates that the human driver assessed less risky behavior as more natural. This is rather unexpected since the human behavior can intentionally or unintentionally show very risky patterns. In further works, we will investigate other features and can compare the scores by showing the same scenes to other drivers for a cross-validation.

Finally, the study showed the relevance of a subjective evaluation of automated vehicle behavior and provides a method to compare automated vehicle algorithm designs and setups. This is especially reasonable when a human like behavior is desired in the automated vehicle control. In our perspective, there are several advantages to aim for this goal since behavior models for the prediction within automated vehicles are commonly designed using normal driving data in a human-driver-only domain. The changes in the behavior that might appear when an automated vehicle is identified remain unconsidered and seem challenging to be integrated. So, an automated vehicle capable of blending in the current traffic environment seems more desirable.

Acknowledgment

This work was supported by Valeo Driving Assistance Research in cooperation with the Valeo research team led by Dr. Benazouz Bradai and project leader Paulo Resende.

References

[1] E. Aria, J. Olstam, and C. Schwietering, “Investigation of Automated Vehicle Effects on Driver’s Behavior and Traffic Performance,” *Transp. Res. Procedia*, vol. 15, pp. 761–770, 2016.

[2] M. Gouy, K. Wiedemann, A. Stevens, G. Brunett, and N. Reed, “Driving next to automated vehicle platoons: How do short time headways influence non-platoon drivers’ longitudinal control?,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 27, pp. 264–273, 2014.

[3] P. De Beaucorps, T. Streubel, A. Verroust-blondet, F. Nashashibi, B. Bradai, and P. Resende, “Decision-making for automated vehicles at intersections adapting human-like behavior,” in *IEEE Intelligent Vehicles Symposium*, 2017.

[4] J. Bärgman, K. Smith, and J. Werneke, “Quantifying drivers’ comfort-zone and dread-zone boundaries in left turn across path/opposite direction (LTAP/OD) scenarios,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 35, pp. 170–184, 2015.

[5] “Real-Time Physics Simulation.” [Online]. Available: bulletphysics.org.

[6] K. Maddock, “Vehicle Simulation with Bullet,” 2010.