

Stopping by looking: A driver-pedestrian interaction study in a coupled simulator using head-mounted displays with eye-tracking

Chun Sang Mok, Pavlo Bazilinskyy, Joost de Winter #

Corresponding author, j.c.f.dewinter@tudelft.nl

December, 2021

Abstract

Automated vehicles (AVs) can perform low-level control tasks but are not always capable of proper decision-making. This paper presents a concept of eye-based maneuver control for AV-pedestrian interaction. Previously, it was unknown whether the AV should conduct a stopping maneuver when the driver looks at the pedestrian or looks away from the pedestrian. A two-agent experiment was conducted using two head-mounted displays with integrated eye-tracking. Seventeen pairs of participants (pedestrian and driver) each interacted in a road crossing scenario. The pedestrians' task was to hold a button when they felt safe to cross the road, and the drivers' task was to direct their gaze according to instructions. Participants completed three 16-trial blocks: (1) Baseline, in which the AV was pre-programmed to yield or not yield, (2) Look to Yield (LTY) in which the AV yielded when the driver looked at the pedestrian, and (3) Look Away to Yield (LATY) in which the AV yielded when the driver did not look at the pedestrian. The driver's eye movements in the LTY and LATY conditions were visualized using a virtual laser. A performance score was computed based on whether the pedestrian held the button when the AV yielded and released the button when the AV did not yield. Furthermore, the pedestrians' and drivers' acceptance of the mappings was measured through a questionnaire. The results showed that the LTY and LATY mappings yielded better crossing performance than Baseline. Furthermore, the LTY condition was best accepted by drivers and pedestrians. Eye-tracking analyses indicated that the LTY and LATY mappings attracted the pedestrian's attention, but pedestrians adequately distributed their attention between the AV and a second vehicle approaching from the other direction. In conclusion, LTY control may be a promising means of AV control at intersections before full automation is technologically feasible.

Keywords: intent communication; virtual reality; eye-tracking; AV-pedestrian interaction; multi-agent interaction

Introduction

Various forms of vehicle automation, such as lane-keeping assistance systems, adaptive cruise control, and autonomous emergency braking, are available on the market today (Bengler et al., 2014). In addition, recently deployed vehicles include traffic light detection, allowing the car to stop in front of an intersection automatically (Nica, 2020; Tesla, 2021). These existing technologies ensure that the car follows a given target (e.g., lane center, headway).

Although current vehicle automation excels in bounded tasks such as those described above, automation technology still has difficulty performing tasks that depend on social context, such as predicting whether a pedestrian will cross the road (Rudenko et al., 2020; Vinkhuyzen & Cefkin, 2016). For this reason, in current AVs, the human driver still supervises the automation, ready to override the system when needed (Lu et al., 2019; Noy et al., 2018). In addition, some vehicles feature a monitoring system that tracks the driver's eye or head movements to verify whether the driver is sufficiently involved (e.g., Cadillac, 2021; Ford, 2021).

One relatively novel way to keep the human involved and allocate responsibility between human and machine is maneuver-based control (Detjen et al., 2021; Hakuli et al., 2011; Manawadu et al., 2018). In maneuver-based control, the driver commands on the maneuver level (e.g., stop the vehicle), while the machine executes tasks on the control level (e.g., braking, turning the steering wheel). Thus, with maneuver-based control, the human is responsible for vital decision-making tasks, whereas low-level control is delegated to automation.

Human-machine interfaces for maneuver-based control typically rely on physical interaction. Examples are a touchscreen (for seminal studies, see Franz et al., 2016; Kauer et al., 2010; Walch et al., 2016), a haptic joystick (Siebert et al., 2019), or a modified steering wheel (Flemisch et al., 2020). In addition, some researchers have proposed touchless interfaces, where the AV is controlled using voice or mid-air gestures (Detjen et al., 2020; Vijayan et al., 2017; Walch et al., 2016).

The present study proposes touchless maneuver-based control using the driver's eye movements. This type of control may be efficient because eye movements coincide with attention (Just & Carpenter, 1980) and because body movement (other than turning the head or eyes) is not required. Outside of the domain of driving, studies have shown that gaze-contingent interfaces can have powerful uses. Examples are eye-typing (Majaranta & R  ih  , 2002), object positioning (Liu et al., 2020) and navigation (Stellmach & Dachsel, 2012) in virtual worlds, training of perceptual skills (Ryu et al., 2016), surveillance displays (Smith et al., 2015), and teleoperation (drone control: Hansen et al., 2014; surgical robot control: Noonan et al., 2008) (see Reingold et al., 2003 and Duchowski, 2018, for reviews of applications of gaze-contingent control and other forms of gaze-based interaction).

So far, only a limited amount of research is available on eye-based control in automated and assisted driving. Ataya et al. (2020) used heads-up gaze-head input, where drivers first gazed at an icon of interest then nodded to confirm their selection and compared this concept with touch-input interaction and voice interaction. They used these input modalities in a driving simulator to obtain drivers' satisfaction ratings of the decision-making of an AV. Jiang et al. (2018) demonstrated a parking assistance system in which drivers indicated the intended direction of their vehicle using eye movements. Wu et al. (2019) developed a system that used the driver's eye gaze to predict the driver's upcoming maneuver. They argued that this concept could be helpful in assistance systems that help prevent dangerous situations. In Wang et al. (2020), drivers requested an assistance system to take action (e.g., change lanes) by looking at an object and uttering "*Watch out!*". Despite these demonstrations of gaze-contingency in driving, there is still little knowledge about how a maneuver-based controlled AV should be controlled by means of eye movements.

When considering a pedestrian-crossing situation, a gaze-controlled AV could adopt two possible mapping strategies. The first strategy, Look to Yield, is that the AV stops when the driver looks at the pedestrian and continues driving when not looking at the pedestrian. This mapping strategy is consistent with research showing that pedestrians are more willing to cross the road when

drivers seek eye contact compared to when the driver does not (Faas et al., 2021; Malmsten Lundgren et al., 2017; Onkhar et al., 2022; Yang, 2017).

The opposite mapping, Look Away to Yield, stops the car when not looking at the pedestrian. Conversely, if the driver is paying attention to the pedestrian, the AV continues to drive. This mapping is loosely consistent with the ‘minimal risk maneuver’ (‘fallback maneuver’) described in automated driving standards (International Organization for Standardization, 2020; Karakaya & Bengler, 2021). More specifically, if the driver is paying attention to potential hazards (pedestrians), the AV can safely continue driving, knowing that the driver is situationally aware. Conversely, if the driver is *not* paying attention to potential hazards, the AV automatically stops out of precaution.

Study Aims and Approach

The study aimed to determine which of the two mappings (Look to Yield, vs. Look Away to Yield) for controlling an AV by means of eye movements is best accepted by pedestrians and drivers. For this purpose, a human-subject study was conducted where the driver and the pedestrian interacted in the same virtual environment. Drivers and pedestrians were immersed using a virtual-reality headset with integrated eye-tracking, and the driver’s eye movements were visualized using a virtual laser. Recent research proposes external human-machine interfaces (eHMLs) to let the AV communicate its intentions to pedestrians (Dey et al., 2020). The laser-like display is a special type of eHML that allowed the driver to confirm that his/her eyes were tracked correctly, whereas it was expected to help the pedestrian predict whether the AV would stop or not.

In real traffic, pedestrians will often have to consider multiple road users simultaneously, something that is often overlooked in research on eHMLs so far (Tabone et al., 2021; for exceptions, see Joisten et al., 2021; Mahadevan et al., 2019; Wilbrink et al., 2021). That is, a risk is that pedestrians will overrely the (attention-grabbing) AV with its laser display but ignore surrounding road users. Therefore, both mapping concepts were evaluated with and without a second vehicle coming from the other direction.

Methods

Participants

Thirty-four participants (17 in the role of driver; 17 in the role of pedestrian) partook in the experiment. The drivers were 13 males and 4 females and had a mean age of 23.6 years ($SD = 2.0$). The pedestrians were 10 males and 7 females and had a mean age of 23.4 years ($SD = 1.8$). The nationalities of the participants were Dutch (28), Chinese (4), and Indian (2). Note that traffic in India is left-handed, but the Indian participants resided in the Netherlands for more than a year.

Five participants (4 drivers, 1 pedestrian) had participated in an experiment on crossing behavior before. Nine of the drivers wore seeing aids (7 glasses and 2 contact lenses), and 10 of the pedestrians wore seeing aids (5 glasses and 5 contact lenses) during the experiment. Moreover, 12 of the drivers and 12 of the pedestrians had a driver’s license, and 12 of the drivers and 11 of the pedestrians had previous experiences with virtual reality glasses. All participants provided written informed consent. The research was approved by the Human Research Ethics Committee of the Delft University of Technology.

Hardware

Two Varjo VR-2 Pro head-mounted displays (HMDs) and two Alienware PCs with identical specifications were used for the experiment. The HMDs were equipped with an industrial-grade,

0.2°-precision eye-tracker and provided a horizontal and vertical field of view of 87° for the participants. The view was provided by two low-persistence micro-OLEDs with a display resolution of 1920x1080 pixels and two low-persistence AMOLEDs with a display resolution of 1440x1600 pixels. The PCs used an Intel Core i9-9900K 3.60 GHz CPU, had 64.0 GB RAM, and used two GPUs: Intel UHD Graphics 630 and NVIDIA GeForce RTX 2081Ti. The pedestrians used an HTC Vive Controller 2.0 to indicate whether they felt safe to cross the road. Two SteamVR base stations were used to track the HMDs and the controller. Finally, a Beyerdynamic DT 770 PRO headset was used by the pedestrian. Data were recorded at a frequency of 60 Hz.

Virtual Reality Environment

The virtual environment was based on the open-source coupled-simulator project of Bazilinskyy et al. (2020). Figure 1 shows the top view of the zebra crossing, including the pedestrian, the AV, and the second vehicle.

The AV approached the zebra crossing from the pedestrian's left side, and the second vehicle approached from the pedestrian's right side. The pedestrian's camera was placed 1.7 m (global average height for men; Roser et al., 2013) above the 0.22-m high curb and at a perpendicular distance of 1.0 m from the edge of the road. The width of the road was 10 m. The pedestrian and driver could rotate their heads in all directions but could not move their body in the virtual world. The driver sat in the passenger seat rather than the driver seat to enhance the impression of being in an AV.

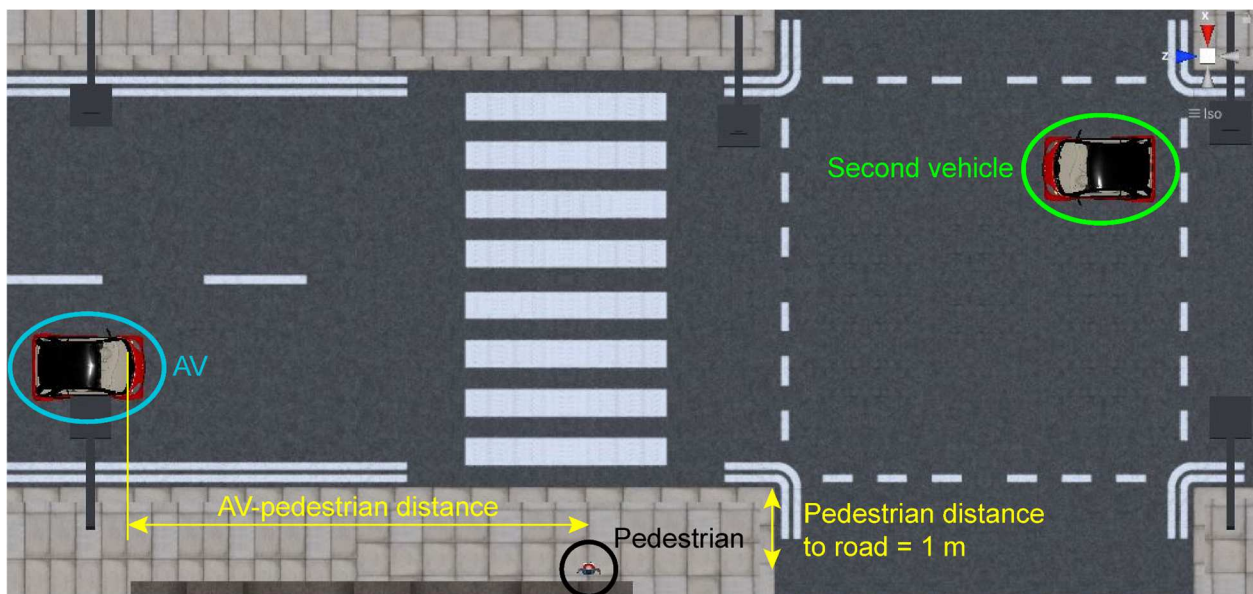


Figure 1. Top view of the virtual environment. In all trials, the AV approached from the left, and the second vehicle approached from the right. The AV-pedestrian distance is defined as the distance along the road measured from the pedestrian to the windshield of the AV.

Eye-Gaze Visualization

The driver's visual attention was rendered as a laser. A 7-cm-wide semi-transparent ($\alpha = 0.2$) laser was drawn between the middle point between the driver's eyes and a point 100 m away in the driver's gaze direction (see Figure 2). The eye-gaze visualization was realized using the head- and eye-tracking capabilities of the Varjo VR-2 Pro HMD.

Experimental Design

Participants were divided into the role of driver or pedestrian before the session began. The experiment was of a within-subject design, with three independent variables: the mapping (Baseline, LTY, LATY), the yielding behavior of the vehicle (yielding, non-yielding), and the presence of a second vehicle (without, with).

The sessions were divided into three blocks. Each block contained one mapping:

- Baseline
- Look to Yield (LTY)
- Look Away to Yield (LATY)

Furthermore, each block contained the same four scenarios:

- Yielding AV – Without second vehicle
- Yielding AV – With second vehicle
- Non-Yielding AV – Without second vehicle
- Non-Yielding AV – With second vehicle

The order of trials within the block was randomized for each participant. All participants began with the Baseline block. This was followed by either the LTY block or the LATY block, depending on the participant number (odd: LTY; even: LATY). In total, the two participants performed 48 trials (3 mappings x 2 yielding scenarios x 2 second-vehicle scenarios x 4 repetitions).

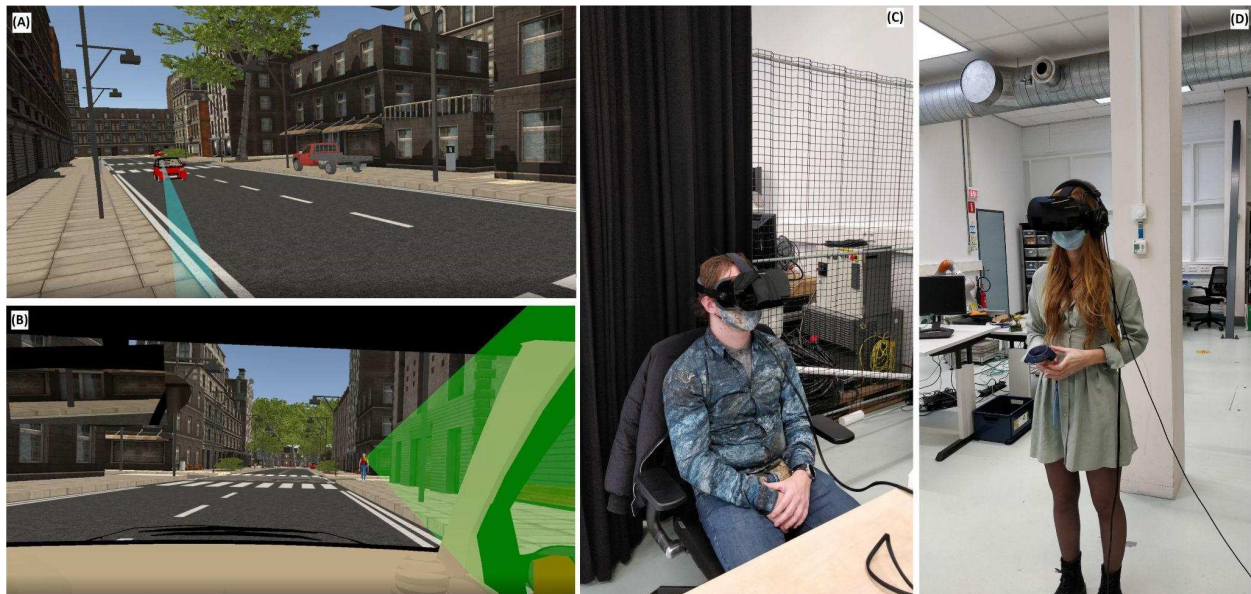


Figure 2. (A) Driver's eye-gaze visualization from the pedestrian's perspective. (B) Driver's eye-gaze visualization from the driver's perspective. (C) Driver wearing the Varjo VR-2 Pro HMD. (D) A pedestrian wearing the Varjo VR-2 Pro HMD and headphones while holding an HTC Vive 2.0 controller.

Vehicle Behavior

The AV's yielding behavior in the Baseline and LATY conditions was not interactive. The driver was instructed to follow the LATY instructions and was therefore expected not to look at the pedestrian. Thus, there would be no manual trigger point for yielding. In real traffic, drivers tend to make their yielding decision later than 30 m before the crosswalk at a vehicle speed of 30 km/h (Schneemann & Gohl, 2016). It was chosen to start the deceleration at an AV-pedestrian distance shorter than 30 m to mimic human deceleration behavior. More specifically, the pre-programmed

braking was initiated at an AV-pedestrian distance of 22.5 m, resulting in the AV coming to a standstill at a distance of 6.2 m from the pedestrian.

For the LTY mapping, the yielding behavior of the AV was interactive. The AV braked when the driver looked at the pedestrian if the AV-pedestrian distance was between 14.4 and 25 m. An adaptive deceleration was used so that the AV would always stop at a distance of 6.2 m from the pedestrian. The yielding trigger needed to be activated before 14.4 m to ensure that the deceleration did not exceed an assumed comfortable deceleration of 3 m/s^2 (Schroeder, 2008). A single glance of the driver in the set distance interval was sufficient to initiate the deceleration of the AV. 'Seeing the pedestrian' was defined as directing the eye gaze at the pedestrian avatar's hitbox (an invisible shape used for real-time collision detection). The rectangle was as tall as the avatar, and the width was 1.9 m, equal to the length of the arms stretched horizontally. At a distance of 25 m, the rectangle width of 1.9 m equals about 4° of the visual field.

The second vehicle came from the right and maintained a constant speed of 30 km/h. In the no-yielding scenario, the second vehicle passed the pedestrian approximately 0.9 s later than the AV, and in the yielding scenario, the second vehicle passed the pedestrian at the same time the AV came to a stop.

Figure 3 shows the AV trajectory of every trial without the second vehicle.

- Phase 1 is the approaching phase and is defined as the period after the start signal till the start of the trigger distance for yielding, i.e., AV-pedestrian distance of 25 m (time between 0 and 4.3 s).
- Phase 2 represents the deceleration phase of the yielding AVs and is defined as the period between an AV-pedestrian distance of 25 m and the vehicle coming to a stop (time between 4.3 and 8.2 s). Note that Phase 2 in LTY shows a variation in time (end time between 7.5 and 8.5 s) for the yielding trials because this condition used adaptive deceleration, as explained above. For non-yielding trials, Phase 2 ends when the vehicle passes the zebra (time between 4.3 and 8.0 s).
- Phase 3 represents the period during which the AV stood still for 2.6 s.

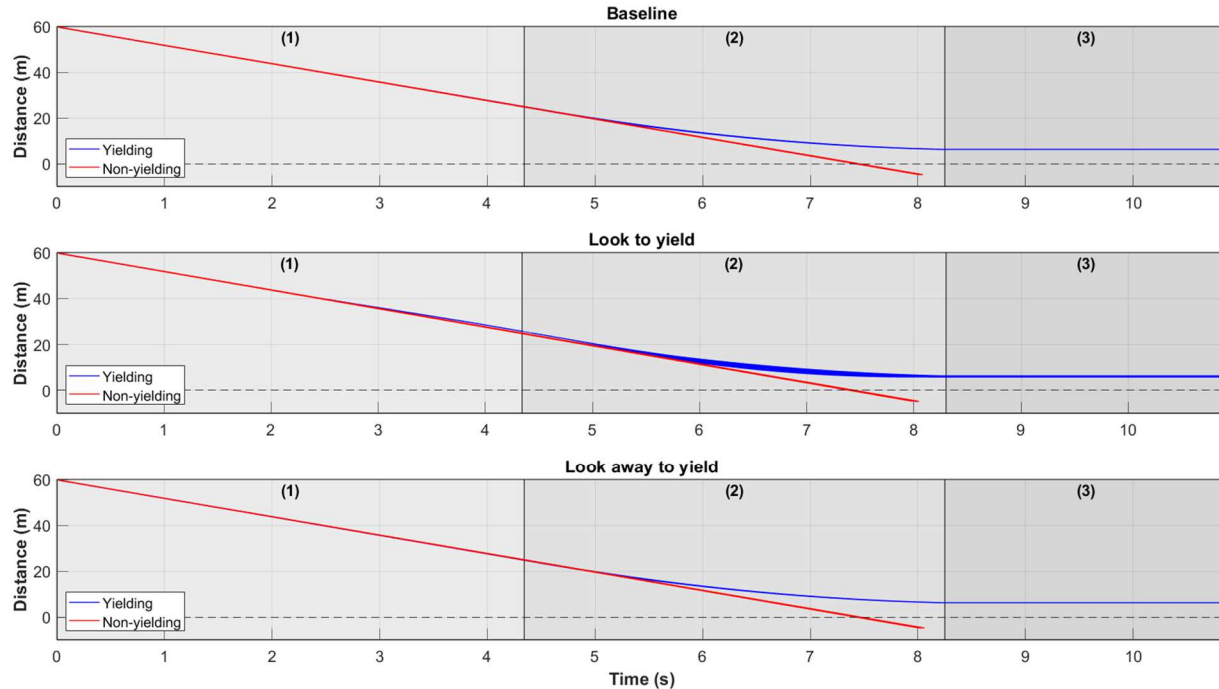


Figure 3. AV-pedestrian distance as a function of time for the Baseline, LTY, and LATY conditions. Phase (1) represents the period after the start signal (0 m) till the start of the yielding trigger detection range (25 m). Phase (2) is the start of the yielding trigger detection range (25 m) till the standstill of the AV, and Phase (3) is the period when the AV was at a standstill for 2.6 s. Note that the figure was created from plotting all trials without a second vehicle (between 64 and 68 trials per condition).

Procedure

Because of COVID-19, participants were requested to disinfect their hands and wear facemasks before entering the lab. Participants were first informed about the purpose of the study and were given an informed consent form to read and sign. Next, the participants read the instructions and completed a pre-experiment questionnaire asking about demographics and prior experience in gaming, VR, driving, and road crossing.

The mapping details (Baseline, LTY, LATY) and participants' tasks were provided in written form. The task for the driver was to follow the instructions given at the beginning of each trial in the form of *"Stop the AV"* and *"Do not stop the AV"*. The pedestrian's task was to press and hold a button when they felt safe to cross the road and release the button when they did not feel safe to cross the road. At the start of the trial, pedestrians heard *"Press now"* via the headphones to ensure that participants had the button pressed at the beginning of the trial. No other sounds were simulated.

Before the start of each session block, the experimenter repeated to the participants what mapping would be used and what was expected from them. The experimenter also mentioned to the pedestrian that the AV carrying the driver would come from the left side of the pedestrian and that the second vehicle would come from the right side.

First, the participants performed a practice run in which one trial per mapping was provided. Then, three blocks of trials were presented to the participants. Between each block, there was a break in which participants were asked to complete a questionnaire regarding their cybersickness state

using a misery scale (MISC; Bos et al., 2005) and acceptance of the system (Van der Laan et al., 1997). In the acceptance questionnaire, participants responded to the statement “*Your judgments of the eye-gaze visualization system are ...*” on nine five-point items. At the end of the experiment, participants were asked to complete a questionnaire about their mapping preference and level of presence (Witmer & Singer, 1998; data not used for this study).

Data Analysis

First, trials in which the driver did not act in accordance with the instructions were excluded. Next, the button-press percentage was computed for Phase 2 per trial (for similar approaches, see De Clercq et al., 2019; Oudshoorn et al., 2021; Sripada et al., 2021). It was chosen to use Phase 2 because differences between mappings were expected to show during this phase. In comparison, in Phase 1, the car is still far away, and in Phase 3, the car is standing still and no difference between mappings is expected.

Acceptance was computed for two subscales: usefulness (5 items) and satisfaction (4 items). The minimum and maximum possible scores were -2 and +2, respectively. Furthermore, a distribution of the driver’s eye-gaze yaw angle was created to illustrate the driver’s gaze behavior. The calculation was done over Phase 2 from the trial. Finally, to illustrate the gaze behavior of the pedestrian during Phase 2 of the trial, a distribution of the yaw angle difference between the pedestrian’s eye-gaze yaw angle and the yaw angle between the AV and the pedestrian was created. When the yaw difference was 0°, the pedestrian looked at the AV. A positive yaw difference means that the pedestrian looked to the left of the AV, and a negative yaw difference means that the pedestrian looked to the right of the AV.

Results

17 out of 544 trials were removed because participants did not comply with the task instructions. The excluded trials corresponded to one of four scenarios:

- In the LTY mapping, the instruction “*Stop the AV*” was given (yielding scenario), but the driver did not look at the pedestrian: 9 trials.
- In the LTY mapping, the instruction “*Do not stop the AV*” was given (non-yielding scenario), but the driver looked at the pedestrian: 2 trials.
- In the LATY mapping, the instruction “*Stop the AV*” was given (yielding scenario), but the driver looked at the pedestrian: 1 trial.
- In the LATY mapping, the instruction “*Do not stop the AV*” was given (non-yielding scenario), but the driver did not look at the pedestrian: 5 trials.

Figure 4 shows the driver’s gaze yaw distribution during the experiment and confirms that the mapping and instructions functioned properly. In the yielding scenarios (left two subfigures), the gaze in the LTY was directed to the right (the pedestrian). The driver’s gaze in the LATY mapping was directed to the left and middle, that is, away from the pedestrian. Oppositely, in the no-yielding scenarios (right two subfigures), the driver’s gaze in the LATY condition was directed to the right, and the gaze in the LTY was directed to the left and middle.

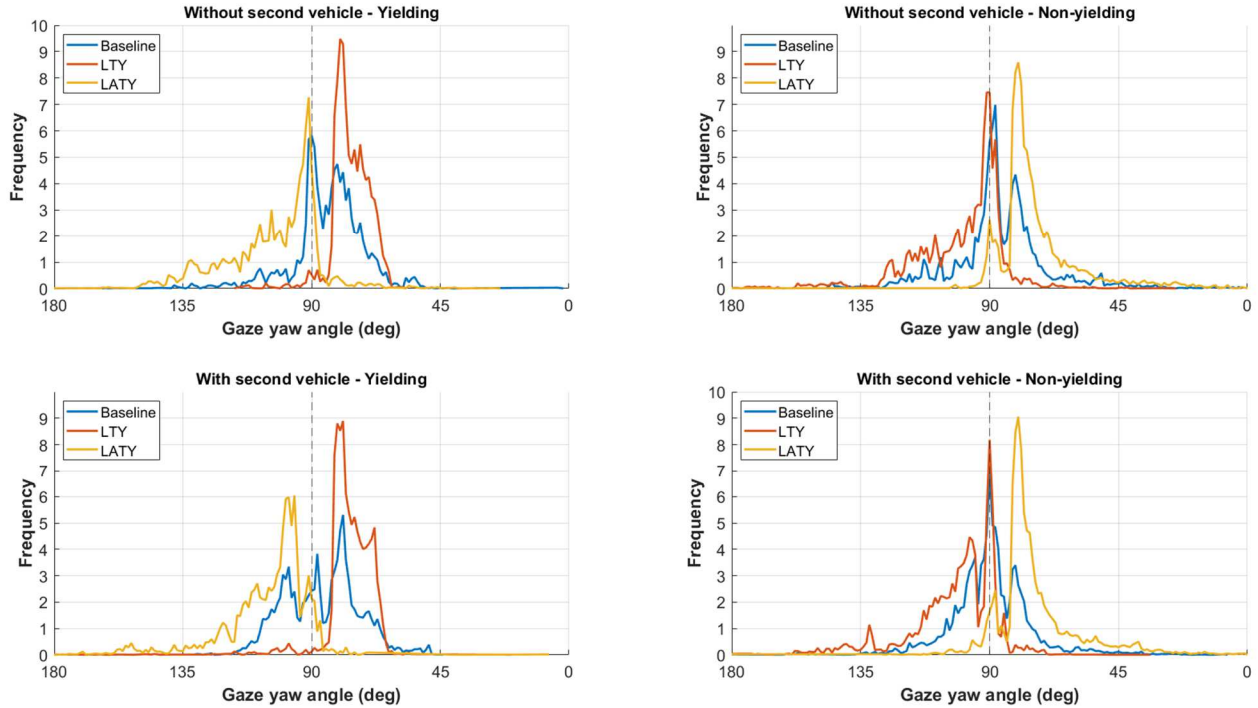


Figure 4. Distribution of the driver's gaze yaw angle in Phase 2. A gaze yaw angle of 90° represents the driver looking straight ahead. A gaze yaw angle smaller than 90° represents the right side, and larger than 90° represents the left side of the driver.

In Figure 5, the button-press data for the scenarios without the second vehicle is shown. A large decrease in button presses was observed in Phase 2 for the yielding scenario of the Baseline condition (solid blue line). This decrease is considerably smaller for the mappings with eye-gaze visualization (red and yellow solid lines). In other words, the LTY and LATY mappings and corresponding eye-gaze visualization made pedestrians believe they could cross when it was safe to cross.

For the non-yielding scenarios, an earlier drop in button presses is found for the LTY and LATY conditions (red and yellow dashed lines) compared to the Baseline condition (blue dashed line). In other words, the LTY and LATY conditions refrained pedestrians from crossing when the AV maintained speed.

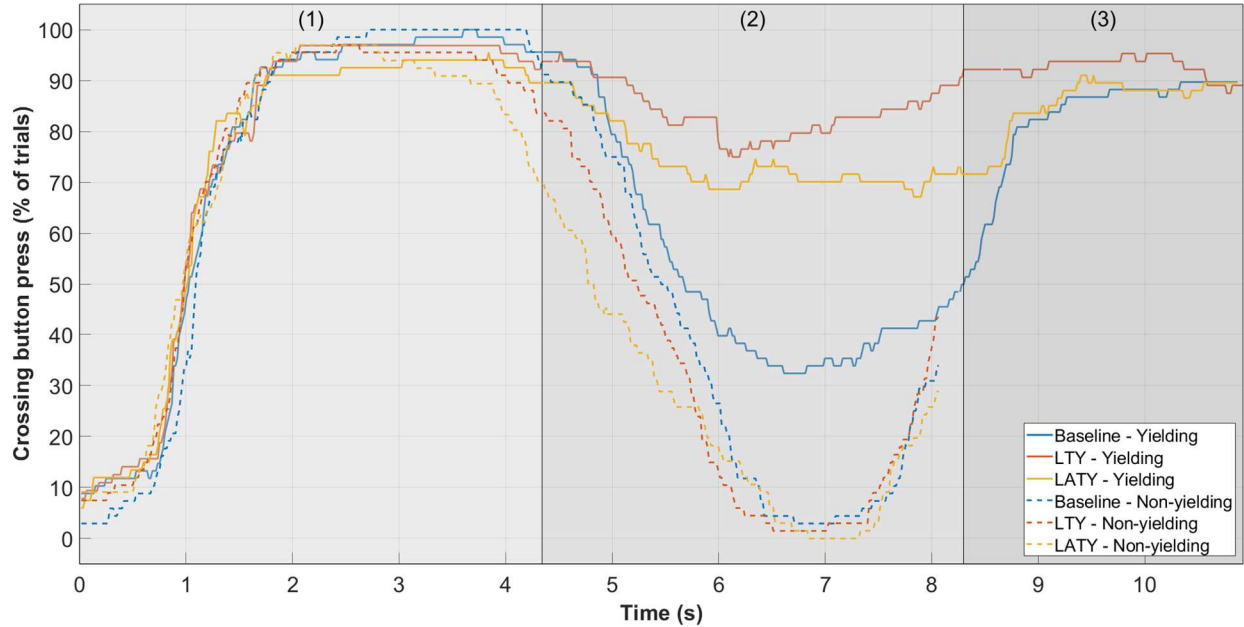


Figure 5. Button-press data for the three mappings for scenarios with yielding and non-yielding AVs, without the presence of the second vehicle.

Figure 6 shows the button-press data for the scenarios with the second vehicle. In the yielding scenario, good performance is characterized by releasing the button in Phase 2 (when the second vehicle is arriving). No clear difference between the three mappings can be seen.

Similar to Figure 5, for the non-yielding scenarios, an earlier drop in button presses is observed for the LTY and LATY mappings (red and yellow dashed lines) compared to Baseline (blue dashed line).

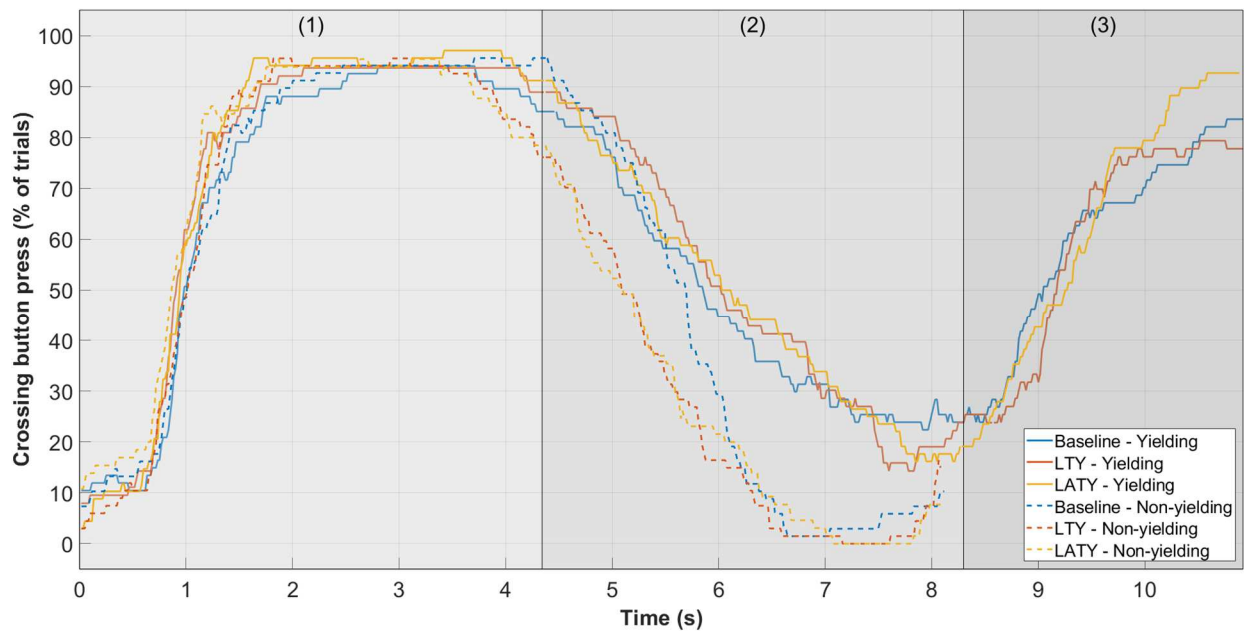


Figure 6. Button-press data for the three mappings for scenarios with yielding and non-yielding AVs, with the presence of the second vehicle.

Table 1 shows the button-press scores for the three mappings. It can be seen that the LTY and LATY mappings yielded a higher score than Baseline when the vehicle was yielding, and a lower score when the vehicle was not yielding. In other words, the LTY and LATY mappings both helped improve performance compared to Baseline. When the second vehicle was present, however, there were no significant differences between the three mappings. In other words, the LTY and LATY mappings did not cause participants to cross when they should not. Finally, no significant difference in button-press scores was observed between the LTY and LATY mappings.

Table 1

Mean (SD) button press score per mapping condition in four different scenarios.

Mapping	1. Without second vehicle, Yielding	2. Without second vehicle, Non-yielding	3. With second vehicle, Yielding	4. With second vehicle, Non-yielding
Baseline	52.3 (30.4)	34.2 (17.1)	43.9 (30.9)	33.5 (16.4)
LTY	84.7 (22.1)	28.1 (16.0)	49.5 (26.8)	22.9 (18.2)
LATY	74.7 (31.5)	22.8 (20.4)	48.0 (28.4)	24.5 (18.6)
Baseline vs. LTY	$t(16) = 5.63$, $p < 0.001$	$t(16) = 1.68$, $p = 0.113$	$t(16) = 1.09$, $p = 0.292$	$t(16) = 3.53$, $p = 0.003$
Baseline vs. LATY	$t(16) = 3.53$, $p = 0.003$	$t(16) = 2.83$, $p = 0.012$	$t(16) = 1.05$, $p = 0.309$	$t(16) = 2.26$, $p = 0.038$
LTY vs. LATY	$t(16) = 1.35$, $p = 0.197$	$t(16) = 1.49$, $p = 0.155$	$t(16) = 0.30$, $p = 0.767$	$t(16) = 0.49$, $p = 0.633$

Figure 7 shows the 'gaze yaw difference' in Phase 2. The peak at 0° represents the pedestrian looking at the AV. For the scenarios without the second vehicle (left two subfigures), LTY (red) has the highest peak when the AV yielded, and LATY has the highest peak when the AV did not yield. These findings suggest that being looked at (yielding & LTY, non-yielding & LATY) encouraged the pedestrian to look back. Figure 7 further shows that pedestrians distributed their attention when the second vehicle was present (right two subfigures), with the peak being about half as high as compared to without the second vehicle (left two subfigures).

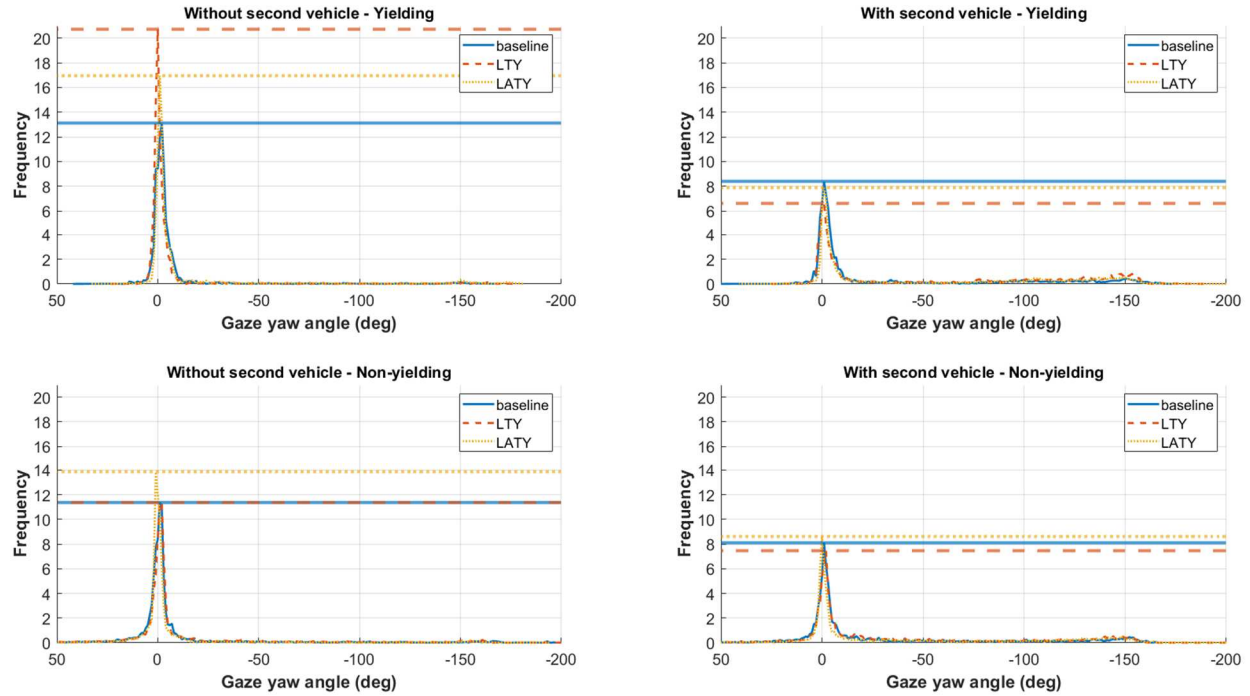


Figure 7. Yaw difference between the AV and pedestrian, per yielding scenario and mapping. The horizontal lines represent the top of the highest peak of the distribution.

Table 2 shows the mean acceptance ratings for the LTY and LATY mappings. The pedestrians experienced the LTY mapping as significantly more satisfactory than the LATY condition. Furthermore, drivers experienced the LTY condition as more useful than the LATY condition, but satisfaction ratings were only moderate, i.e., near the midpoint of the scale.

These observations were supported by a post-experiment questionnaire in which participants were asked to rank the three mappings (Baseline, LTY, LATY) in terms of their preference. The results revealed the following:

- 10 of 17 drivers ranked the LTY mapping as the most preferred, and 4 ranked it as the least preferred (i.e., third ranking).
- 3 of 17 drivers ranked the LATY condition as the most preferred, and 10 ranked it as the least preferred.
- 14 of 17 pedestrians ranked the LTY condition as the most preferred, and 0 ranked it as the least preferred.
- 0 of 17 pedestrians ranked the LATY condition as the most preferred, and 8 ranked it as the least preferred.

Additionally, in the post-block questionnaire, pedestrians were asked whether it was clear to them that the vehicle was going to yield. The mean was 5.65 for LTY, and 5.00 for LATY, on a scale of 1 = Strongly disagree to 7 = Strongly agree. Drivers were asked whether it was easy for them to direct the eye-gaze visualization to where they wanted it to be. The mean was 4.71 for LTY and 5.53 for LATY on a scale of 1 = Extremely difficult to 7 = Extremely easy.

Table 2

Mean (SD) acceptance scores for pedestrians ($n = 17$) and drivers ($n = 17$).

Mapping	Usefulness, pedestrians	Satisfaction, pedestrians	Usefulness, drivers	Satisfaction, drivers
LTY	1.00 (0.61)	0.96 (0.55)	0.93 (0.62)	0.01 (1.02)
LATY	0.82 (0.73)	0.65 (0.62)	0.47 (0.82)	-0.18 (0.95)
LTY vs. LATY	$t(16) = 2.06$, $p = 0.056$	$t(16) = 3.92$, $p = 0.001$	$t(16) = 2.19$, $p = 0.043$	$t(16) = 0.91$, $p = 0.375$

Discussion

This study compared two mappings for maneuver-based AV control combined with driver eye-movement visualization. In the LTU mapping, the AV stopped for the pedestrian when the driver looked at the pedestrian. Conversely, in the LATY mapping, the AV stopped when the driver did not look at the pedestrian. An experiment was conducted with 34 participants, divided into 17 pairs of two, present in the same virtual environment. Thus, participants were a driver of a maneuver-based controlled AV or they were a pedestrian who had to indicate their intentions to cross the road via a button-press.

Pressing the button when the AV was yielding and not pressing when the vehicle was not yielding were regarded as indicative of high performance. However, when a second vehicle was present, the pedestrian had to refrain from crossing the road; hence in this situation, a low button-press rate was regarded as indicative of high performance. The scenario with the second vehicle was included to test for pedestrian overreliance on the driver's eye-gaze visualization.

The pedestrians' button-press results indicated that the LTU and LATY mappings both yielded improved performance compared to the Baseline condition without eye-gaze visualization. These findings are consistent with much literature showing that eHMLs provide clarity to pedestrians (e.g., Böckle et al., 2017; Chang et al., 2017; De Clercq et al., 2019; Oudshoorn et al., 2021). The eye-gaze visualization can be seen as a suitable eHML because, firstly, it depicts the driver's intention without further instruction or clarification, consistent with recommendations for eHMLs outlined by the International Organization for Standardization (2018; see also Tabone et al., 2021; Benderius et al., 2018). Second, the literature suggests that non-directional signaling by eHMLs could confuse pedestrians as it may be unclear from whom the message is meant (Dey et al., 2021; Hensch et al., 2019). Through eye-gaze visualization, the message can be directed towards one specific pedestrian. Lastly, in real traffic, it may be unclear whether drivers and pedestrians have seen each other due to sunlight or glare, for example (AlAdawy et al., 2019). The visualization of eye gaze makes such information salient and explicit. The current study compared the concepts for maneuver-based control with a Baseline condition. A comparison with existing eHMLs would be a welcome topic of future research (for one study on directed vs. undirected signaling, see Dietrich et al., 2018).

The results showed no significant differences in pedestrians' crossing intentions between the LTU and LATY mappings, which may be explained because drivers and pedestrians were instructed about the meaning of the mappings. The LTU and LATY mappings provided the same information to pedestrians (only in opposite directions), and hence no difference in crossing performance ought to be expected. Of note, the LTU and LATY mappings, which informed the pedestrian about whether they could cross, did not cause overreliance in the presence of a second vehicle approaching from the other direction. Although the eye-movement analysis showed that the LTU

and LATY mappings attracted the pedestrians' attention, pedestrians distributed their attention between the AV and second vehicle.

Although no significant differences in crossing performance between the LTY and LATY conditions were observed, there were notable differences in acceptance. The LTY condition was regarded as most satisfactory by pedestrians, and most useful by drivers. Furthermore, when asked to rank the mappings in terms of preference, participants ranked the LTY mapping the highest. Research shows that drivers are inclined to look at pedestrians and other potential hazards (e.g., Underwood et al., 2011); hence, deliberately not looking at a pedestrian may be regarded as counterintuitive. One driver wrote in the post-block questionnaire: *"I had to think hard about if I had to look away or not. It was not logical to me."* Similarly, recent research indicates that pedestrians are more likely to cross the road when the driver looks at them (Faas et al., 2021; Malmsten Lundgren et al., 2017; Onkhar et al., 2022). One pedestrian stated: *"This map [LATY] was confusing me a bit. When the spotlights (the lasers) were on me, I was thinking that it was my turn to cross the road, while I actually should wait."*

A noteworthy aspect of the experiment is the use of two participants in the same virtual world, an approach that is still relatively rare but gaining popularity in human factors research (Hancock & De Ridder, 2003; Houtenbos et al., 2017; Ko et al., 2022; Lehsing et al., 2015; Muehlbacher et al., 2014; Oeltze & Schießl, 2015; S. Park et al., 2019; Preuk et al., 2016). The visualization of a human driver's eye-gaze may have contributed to more natural and realistic situations compared to a pre-programmed laser beam. However, it does introduce more variation as every person behaves somewhat differently. Another noteworthy aspect of the current research is the combination of eye-tracking and head tracking in an HMD, something that is gaining popularity (Blattgerste et al., 2018; Chen et al., 2018; Ferris et al., 2018; J. H. Park et al., 2019) but still relatively rare. Head-tracking in the HMD allowed the pedestrian to look around, resulting in a high field of regard. The wide eye-gaze distribution shown in the results section would be impossible to obtain using computer monitors or driving simulators.

A limitation of this research is that the participants lacked diversity in terms of age and nationality. The increase of age contributes to cognitive and sensory decline, resulting in older people taking fewer risks (e.g., Dunbar et al., 2004). Nationality may influence crossing behavior due to differences in cultural and social norms across countries (Pelé et al., 2017; Ranasinghe et al., 2020), although recent research suggests that the effects of eHMLs on pedestrians' crossing intentions are cross-nationally robust (Oudshoorn et al., 2021).

Like many other AV-pedestrian interaction studies, the experiment addressed a single-pedestrian crossing in front of one or two vehicles. In reality, however, a pedestrian crossing situation is more complex. In (future) traffic, pedestrians may have to deal with a large number of other pedestrians and AVs at different levels of automation, as well as different crossing configurations and different weather conditions (for recommendations, see Dey et al., 2020). Furthermore, even with a high level of presence in virtual reality, it is still unclear whether testing in virtual environments is as valid as naturalistic testing (Tabone et al., 2021).

Regarding the experiment, the block order randomization did not include the Baseline block as it was expected that there would be only minimal learning effects due to the addition of a practice run. Our analysis of learning curves showed no major learning effects, with the Baseline mapping yielding consistently lower performance (i.e., lower button-press rates for the yielding scenario without second vehicle, higher button-press rates for non-yielding scenarios) than the LTY and LATY mappings (see Figure S1 in the supplementary material).

The technical and practical feasibility of the eye-gaze visualizations were not considered in the design process. The visualization of the driver's eye gaze brings many challenges. Eye-trackers and a laser projector system would need to be installed in the AV. Moreover, the eye-gaze visualization could distract other drivers and potentially lead to accidents. One also needs to consider the visibility of the eye-gaze visualization in changing weather conditions. Another approach to implementing eye-gaze visualization is augmented reality (AR). AR would solve the problem of distracting other road users and visibility in changing weather conditions as the visualization is displayed directly on the wearable. However, many challenges regarding augmented reality need to be addressed in future traffic, such as privacy, invasiveness, user-friendliness, technological feasibility, and inclusiveness (Tabone et al., 2021).

Conclusions

This study compared two forms of eye-based AV control combined with the driver's eye-gaze visualization. The LTY and LATY conditions improved the crossing interaction compared to the Baseline condition. Furthermore, the LTY mapping achieved higher acceptance ratings than the LATY condition. This study provides insights that may be important for the future control of automated vehicles, particularly the control of the vehicle through eye movements.

Supplementary Material

Open-source code of the simulator: <https://github.com/bazilinskyy/coupled-sim>

Demo video of the LTY and LATY mappings can be found here:

<https://www.dropbox.com/sh/xuoq7k8w563boj5/AACXaf2jYSJwF0lkGalm2FSDa?dl=0>

References

- AlAdawy, D., Glazer, M., Terwilliger, J., Schmidt, H., Domeyer, J., Mehler, B., Reimer, B., & Fridman, L. (2019). Eye contact between pedestrians and drivers. *Proceedings of the Tenth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Santa Fe, New Mexico, 301–307.
https://drivingassessment.uiowa.edu/sites/drivingassessment.uiowa.edu/files/da2019_47_aladawy_final.pdf
- Ataya, A., Kim, W., Elsharkawy, A., & Kim, S. (2020). Gaze-head input: Examining potential interaction with immediate experience sampling in an autonomous vehicle. *Applied Sciences*, 10, 9011. <https://doi.org/10.3390/app10249011>
- Bazilinskyy, P., Kooijman, L., Dodou, D., & De Winter, J. C. F. (2020). Coupled simulator for research on the interaction between pedestrians and (automated) vehicles. *Proceedings of the 19th Driving Simulation Conference Europe*, Antibes, France.
- Benderius, O., Berger, C., & Lundgren, V. M. (2018). The best rated human–machine interface design for autonomous vehicles in the 2016 grand cooperative driving challenge. *IEEE Transactions on Intelligent Transportation Systems*, 19, 1302–1307.
<https://doi.org/10.1109/TITS.2017.2749970>
- Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine*, 6, 6–22. <https://doi.org/10.1109/MITS.2014.2336271>
- Blattgerste, J., Renner, P., & Pfeiffer, T. (2018). Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. *Proceedings of the Workshop on Communication by Gaze Interaction*, Warsaw, Poland.
<https://doi.org/10.1145/3206343.3206349>
- Böckle, M. P., Brenden, A. P., Klingegård, M., Habibovic, A., & Bout, M. (2017). SAV2P—Exploring the impact of an interface for shared automated vehicles on pedestrians' experience. *Proceedings of the 9th International ACM Conference on Automotive User*

Interfaces and Interactive Vehicular Applications, Oldenburg, Germany.

<https://doi.org/10.1145/3131726.3131765>

- Bos, J. E., MacKinnon, S. N., & Patterson, A. (2005). Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view. *Aviation, Space, and Environmental Medicine*, 76, 1111–1118.
- Cadillac. (2021). 2021 Cadillac Escalade owner's manual. <https://my.cadillac.com/content/dam/gmownercenter/gmna/dynamic/manuals/2021/cadillac/escalade/2021-cadillac-escalade-owners-manual.pdf>
- Chang, C. M., Toda, K., Sakamoto, D., & Igarashi, T. (2017). Eyes on a car: an interface design for communication between an autonomous car and a pedestrian. *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Oldenburg, Germany, 65–73. <https://doi.org/10.1145/3122986.3122989>
- Chen, S. Y., Gao, L., Lai, Y. K., Rosin, P. L., & Xia, S. (2018). Real-time 3D face reconstruction and gaze tracking for virtual reality. *Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces*, Tuebingen/Reutlingen, Germany, 525–526. <https://doi.org/10.1109/VR.2018.8446494>
- De Clercq, G. K., Dietrich, A., Núñez Velasco, P., De Winter, J. C. F., & Happee, R. (2019). External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. *Human Factors*, 61, 1353–1370. <https://doi.org/10.1177/0018720819836343>
- Detjen, H., Geisler, S., & Schneegass, S. (2020). Maneuver-based control interventions during automated driving: Comparing touch, voice, and mid-air gestures as input modalities. *Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics*, Toronto, Canada, 3268–3274. <https://doi.org/10.1109/SMC42975.2020.9283431>
- Detjen, H., Geisler, S., & Schneegass, S. (2021). Driving as side task: Exploring intuitive input modalities for multitasking in automated vehicles. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan. <https://doi.org/10.1145/3411763.3451803>
- Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Rienner, A., Martens, M., & Terken, J. (2020). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. *Transportation Research Interdisciplinary Perspectives*, 7, 100174. <https://doi.org/10.1016/j.trip.2020.100174>
- Dey, D., Van Vastenhoven, A., Cuijpers, R. H., Martens, M., & Pfleging, B. (2021). Towards Scalable eHMIs: Designing for AV-VRU communication beyond one pedestrian. *Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Leeds, UK, 274–286. <https://doi.org/10.1145/3409118.3475129>
- Dietrich, A., Willrodt, J.-H., Wagner, K., & Bengler, K. (2018). Projection-based external human-machine interfaces – Enabling interaction between automated vehicles and pedestrians. *Proceedings of the 17th Driving Simulation Conference Europe*, Antibes, France, 43–50.
- Dunbar, G., Holland, C., & Maylor, E. A. (2004). *Older pedestrians: A critical review of the literature* (Road Safety Research Report No. 37). London: Department for Transport.
- Faas, S. M., Stange, V., & Baumann, M. (2021). Self-driving vehicles and pedestrian interaction: does an external human-machine interface mitigate the threat of a tinted windshield or a distracted driver? *International Journal of Human-Computer Interaction*, 37, 1364–1374. <https://doi.org/10.1080/10447318.2021.1886483>
- Ferris, L., Bond, R., McNeice, L., Grimley, A., Taylor, A., Magee, J., Lyons, F., & Charles, D. (2018). Virtual reality simulation and eye tracking to assess hazard perception of car drivers. *Proceedings of the 32nd International BCS Human Computer Interaction Conference*, Belfast, UK. <https://doi.org/10.14236/ewic/HCI2018.189>
- Flemisch, F., Diederichs, F., Meyer, R., Herzberger, N., Baier, R., Altendorf, E., Spies, J., Usai, M., Kaim, V., Doentgen, B., Bopp-Bertenbreiter, A. V., Widloither, H., Ruth-Schumacher, S.,

- Arzt, C., & Kaiser, F. (2020). Vorreiter: manoeuvre-based steering gestures for partially and highly automated driving. In G. Meixner (Ed.), *Smart automotive mobility* (pp. 231–304). Cham: Springer. https://doi.org/10.1007/978-3-030-45131-8_5
- Ford. (2021). Ford Co-Pilot360™ technology. <https://www.ford.com/technology/driver-assist-technology>
- Franz, B., Kauer, M., Geyer, S., & Hakuli, S. (2016). Conduct-by-Wire. In H. Winner, S. Hakuli, F. Lotz, & C. Singer (Eds.), *Handbook of driver assistance systems* (pp. 1483–1497). Cham: Springer. https://doi.org/10.1007/978-3-319-12352-3_59
- Hakuli, S., Geyer, S., Winner, H., & Henning, J. (2011). Integrated concept for maneuver-based driver assistance. *ATZ Worldwide eMagazine*, 113, 40–42. <https://doi.org/10.1365/s38311-011-0032-5>
- Hancock, P. A., & De Ridder, S. N. (2003). Behavioural accident avoidance science: understanding response in collision incipient conditions. *Ergonomics*, 46, 1111–1135. <https://doi.org/10.1080/0014013031000136386>
- Hansen, J. P., Alapetite, A., MacKenzie, I. S., & Møllenbach, E. (2014). The use of gaze to control drones. *Proceedings of the Symposium on Eye Tracking Research and Applications*, Safety Harbor, FL, 27–34. <https://doi.org/10.1145/2578153.2578156>
- Hensch, A. C., Neumann, I., Beggiato, M., Halama, J., & Krems, J. F. (2019). How should automated vehicles communicate?—Effects of a light-based communication approach in a Wizard-of-Oz study. In N. Stanton (Ed.), *Advances in Human Factors of Transportation* (pp. 79–91). Cham: Springer. https://doi.org/10.1007/978-3-030-20503-4_8
- Houtenbos, M. D., De Winter, J. C. F., Hale, A. R., Wieringa, P. A., & Hagenzieker, M. P. (2017). Concurrent audio-visual feedback for supporting drivers at intersections: A study using two linked driving simulators. *Applied Ergonomics*, 60, 30–42. <https://doi.org/10.1016/j.apergo.2016.10.010>
- International Organization for Standardization. (2018). ISO/TR 23049:2018 Road Vehicles — Ergonomic aspects of external visual communication from automated vehicles to other road users. <https://www.iso.org/standard/74397.html>
- International Organization for Standardization. (2020). ISO/TR 4804:2020 Road vehicles — Safety and cybersecurity for automated driving systems — Design, verification and validation. <https://www.iso.org/standard/80363.html>
- Jiang, Y. S., Warnell, G., Munera, E., & Stone, P. (2018). A study of human-robot copilot systems for en-route destination changing. *Proceedings of the 2018 27th IEEE International Symposium on Robot and Human Interactive Communication*, Nanjing, China, 997–1004. <https://doi.org/10.1109/ROMAN.2018.8525701>
- Joisten, P., Liu, Z., Theobald, N., Webler, A., & Abendroth, B. (2021). Communication of automated vehicles and pedestrian groups: An intercultural study on pedestrians' street crossing decisions. *Proceedings of Mensch und Computer 2021*, Ingolstadt, Germany, 49–53. <https://doi.org/10.1145/3473856.3474004>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Karakaya, B., & Bengler, K. (2021). Investigation of driver behavior during minimal risk maneuvers of automated vehicles. In N. L. Black, W. P. Neumann, & I. Noy (Eds.), *Proceedings of the 21st Congress of the International Ergonomics Association* (pp. 691–700). Cham: Springer. https://doi.org/10.1007/978-3-030-74608-7_84
- Kauer, M., Schreiber, M., & Bruder, R. (2010). How to conduct a car? A design example for maneuver based driver-vehicle interaction. *Proceedings of the 2010 IEEE Intelligent Vehicles Symposium*, La Jolla, CA, 1214–1221. <https://doi.org/10.1109/IVS.2010.5548099>

- Ko, J., Jang, J., & Oh, C. (2022). A multi-agent driving simulation approach for evaluating the safety benefits of connected vehicles. *IEEE Transactions on Intelligent Transportation Systems*. <https://doi.org/10.1109/TITS.2020.3045675>
- Lehsing, C., Kracke, A., & Bengler, K. (2015). Urban perception – A cross-correlation approach to quantify the social interaction in a multiple simulator setting. *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Gran Canaria, Spain, 1014–1021. <https://doi.org/10.1109/ITSC.2015.169>
- Liu, C., Plopski, A., & Orlosky, J. (2020). OrthoGaze: Gaze-based three-dimensional object manipulation using orthogonal planes. *Computers & Graphics*, 89, 1–10. <https://doi.org/10.1016/j.cag.2020.04.005>
- Lu, Z., Zhang, B., Feldhütter, A., Happee, R., Martens, M., & De Winter, J. C. F. (2019). Beyond mere take-over requests: The effects of monitoring requests on driver attention, take-over performance, and acceptance. *Transportation Research Part F: Traffic Psychology and Behaviour*, 63, 22–37. <https://doi.org/10.1016/j.trf.2019.03.018>
- Mahadevan, K., Sanoubari, E., Somanath, S., Young, J. E., & Sharlin, E. (2019). AV-pedestrian interaction design using a pedestrian mixed traffic simulator. *Proceedings of the 2019 on Designing Interactive Systems Conference*, San Diego, CA, 475–486. <https://doi.org/10.1145/3322276.3322328>
- Majaranta, P., & Rähkä, K. J. (2002). Twenty years of eye typing: Systems and design issues. *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, New Orleans, LA, 15–22. <https://doi.org/10.1145/507072.507076>
- Malmsten Lundgren, V., Habibovic, A., Andersson, J., Lagström, T., Nilsson, M., Sirkka, A., Fagerlön, J., Fredriksson, R., Edgren, C., Krupenia, S., & Saluäär, D. (2017). Will there be new communication needs when introducing automated vehicles to the urban context? In N. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.), *Advances in human aspects of transportation. Advances in intelligent systems and computing* (pp. 485–497). Cham: Springer. https://doi.org/10.1007/978-3-319-41682-3_41
- Manawadu, U. E., Hayashi, H., Kamezaki, M., & Sugano, S. (2018). Human-agent collaborative control in automated vehicles for takeover situations in dynamic unstructured urban environments. *IEEE International Conference on Robotics and Automation*, Brisbane, Australia.
- Muehlbacher, D., Rittger, L., & Maag, C. (2014). Real vs. simulated surrounding traffic—Does it matter? *Proceedings of the Driving Simulation Conference Europe*, Paris, France, 22.1–22.6.
- Nica, G. (2020). New BMW Urban Cruise Control recognizes traffic lights. <https://www.bmwblog.com/2020/09/24/new-bmw-urban-cruise-control>
- Noonan, D. P., Mylonas, G. P., Darzi, A., & Yang, G. Z. (2008). Gaze contingent articulated robot control for robot assisted minimally invasive surgery. *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France, 1186–1191. <https://doi.org/10.1109/IROS.2008.4651105>
- Noy, I. Y., Shinar, D., & Horrey, W. J. (2018). Automated driving: Safety blind spots. *Safety Science*, 102, 68–78. <https://doi.org/10.1016/j.ssci.2017.07.018>
- Oeltze, K., & Schiebl, C. (2015). Benefits and challenges of multi-driver simulator studies. *IET Intelligent Transport Systems*, 9, 618–625. <https://doi.org/10.1049/iet-its.2014.0210>
- Onkhar, V., Bazilinskyy, P., Dodou, D., & De Winter, J. C. F. (2022). The effect of drivers' eye contact on pedestrians' perceived safety. *Transportation Research Part F: Traffic Psychology and Behaviour*, 84, 194–210. <https://doi.org/10.1016/j.trf.2021.10.017>
- Oudshoorn, M., De Winter, J. C. F., Bazilinskyy, P., & Dodou, D. (2021). Bio-inspired intent communication for automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, 80, 127–140. <https://doi.org/10.1016/j.trf.2021.03.021>
- Park, J. H., Jeon, H. J., Lim, E.-C., Koo, J.-W., Lee, H. J., Kim, H.-J., Lee, J. S., Song, C.-G., & Hong, S. K. (2019). Feasibility of eye tracking assisted vestibular rehabilitation strategy using

- immersive virtual reality. *Clinical and Experimental Otorhinolaryngology*, 12, 376–384.
<https://doi.org/10.21053/ceo.2018.01592>
- Park, S., Oh, C., Kim, Y., Choi, S., & Park, S. (2019). Understanding impacts of aggressive driving on freeway safety and mobility: A multi-agent driving simulation approach. *Transportation Research Part F: Traffic Psychology and Behaviour*, 64, 377–387.
<https://doi.org/10.1016/j.trf.2019.05.017>
- Pelé, M., Bellut, C., Debergue, E., Gauvin, C., Jeanneret, A., Leclere, T., Nicolas, L., Pontier, F., Zausa, D., & Sueur, C. (2017). Cultural influence of social information use in pedestrian road-crossing behaviours. *Royal Society Open Science*, 4, 160739.
<https://doi.org/10.1098/rsos.160739>
- Preuk, K., Stemmler, E., Schiebl, C., & Jipp, M. (2016). Does assisted driving behavior lead to safety-critical encounters with unequipped vehicles' drivers? *Accident Analysis & Prevention*, 95, 149–156. <https://doi.org/10.1016/j.aap.2016.07.003>
- Ranasinghe, C., Holländer, K., Currano, R., Sirkin, D., Moore, D., Schneegass, S., & Ju, W. (2020). Autonomous vehicle-pedestrian interaction across cultures: Towards designing better external Human Machine Interfaces (eHMI). *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI.
<https://doi.org/10.1145/3334480.3382957>
- Reingold, E. M., Loschky, L. C., McConkie, G. W., & Stampe, D. M. (2003). Gaze-contingent multiresolutional displays: An integrative review. *Human Factors*, 45, 307–328.
<https://doi.org/10.1518/hfes.45.2.307.27235>
- Roser, M., Appel, C., & Ritchie, H. (2013). Human height. <https://ourworldindata.org/human-height>
- Rudenko, A., Palmieri, L., Herman, M., Kitani, K. M., Gavrilu, D. M., & Arras, K. O. (2020). Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, 39, 895–935. <https://doi.org/10.1177/0278364920917446>
- Ryu, D., Mann, D. L., Abernethy, B., & Poolton, J. M. (2016). Gaze-contingent training enhances perceptual skill acquisition. *Journal of Vision*, 16. <https://doi.org/10.1167/16.2.2>
- Schneemann, F., & Gohl, I. (2016). Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. *2016 IEEE Intelligent Vehicles Symposium (IV)*, 38–43. <https://doi.org/10.1109/IVS.2016.7535361>
- Schroeder, B. (2008). *A behavior-based methodology for evaluating pedestrian-vehicle interaction at crosswalk* (PhD dissertation). North Carolina State University.
- Siebert, F. W., Radtke, F., Kiyonaga, E., & Höger, R. (2019). Adjustable automation and manoeuvre control in automated driving. *IET Intelligent Transport Systems*, 13, 1780–1784.
<https://doi.org/10.1049/iet-its.2018.5471>
- Smith, M. A., Mead, P. R., Squire, P. N., Coons, R. L., & Mead, A. S. (2015). Look, don't touch: Evaluation of a gaze-based interface control system. *Ergonomics in Design*, 23, 10–15.
<https://doi.org/10.1177/1064804613516113>
- Sripada, A., Bazilinskyy, P., & De Winter, J. C. F. (2021). Automated vehicles that communicate implicitly: Examining the use of lateral position within the lane. *Ergonomics*, 64, 1416–1428.
<https://doi.org/10.1080/00140139.2021.1925353>
- Stellmach, S., & Dachsel, R. (2012). Designing gaze-based user interfaces for steering in virtual environments. *Proceedings of the Symposium on Eye Tracking Research and Applications*, Santa Barbara, CA, 131–138. <https://doi.org/10.1145/2168556.2168577>
- Tabone, W., De Winter, J., Ackermann, C., Bärghman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D., Sheridan, T. B., & Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives*, 9, 100293.
<https://doi.org/10.1016/j.trip.2020.100293>

- Tesla. (2021). Traffic Light and Stop Sign Control.
https://www.tesla.com/ownersmanual/modely/en_eu/GUID-A701F7DC-875C-4491-BC84-605A77EA152C.html
- Underwood, G., Crundall, D., & Chapman, P. (2011). Driving simulator validation with hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14, 435–446.
<https://doi.org/10.1016/j.trf.2011.04.008>
- Van der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, 5, 1–10. [https://doi.org/10.1016/S0968-090X\(96\)00025-3](https://doi.org/10.1016/S0968-090X(96)00025-3)
- Vijayan, I., Laur, M. H., & Absmeier, J. P. (2017). *Automated vehicle operation based on gesture to pedestrian* (Patent Application No. WO2017119964A1). World Intellectual Property Organization.
- Vinkhuyzen, E., & Cefkin, M. (2016). Developing socially acceptable autonomous vehicles. *Ethnographic Praxis in Industry Conference Proceedings*, 2016, 522–534.
<https://doi.org/10.1111/1559-8918.2016.01108>
- Walch, M., Sieber, T., Hock, P., Baumann, M., & Weber, M. (2016). Towards cooperative driving: involving the driver in an autonomous vehicle's decision making. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Ann Arbor, MI, 261–268. <https://doi.org/10.1145/3003715.3005458>
- Wang, C., Krüger, M., & Wiebel-Herboth, C. B. (2020). “Watch out!”: Prediction-level intervention for automated driving. *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. New York, NY, 169–180.
<https://doi.org/10.1145/3409120.3410652>
- Wilbrink, M., Nuttelmann, M., & Oehl, M. (2021). Scaling up automated vehicles’ eHMI communication designs to interactions with multiple pedestrians—putting eHMIs to the test. *Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Leeds, UK, 119–122.
<https://doi.org/10.1145/3473682.3480277>
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7, 225–240. <https://doi.org/10.1162/105474698565686>
- Wu, M., Louw, T., Lahijanian, M., Ruan, W., Huang, X., Merat, N., & Kwiakowska, M. (2019). Gaze-based intention anticipation over driving manoeuvres in semi-autonomous vehicles. *Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Macau, China, 6210–6216. <https://doi.org/10.1109/IROS40897.2019.8967779>
- Yang, S. (2017). *Driver behavior impact on pedestrians’ crossing experience in the conditionally autonomous driving context* (Doctoral dissertation). KTH Royal Institute of Technology.
<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-220545>

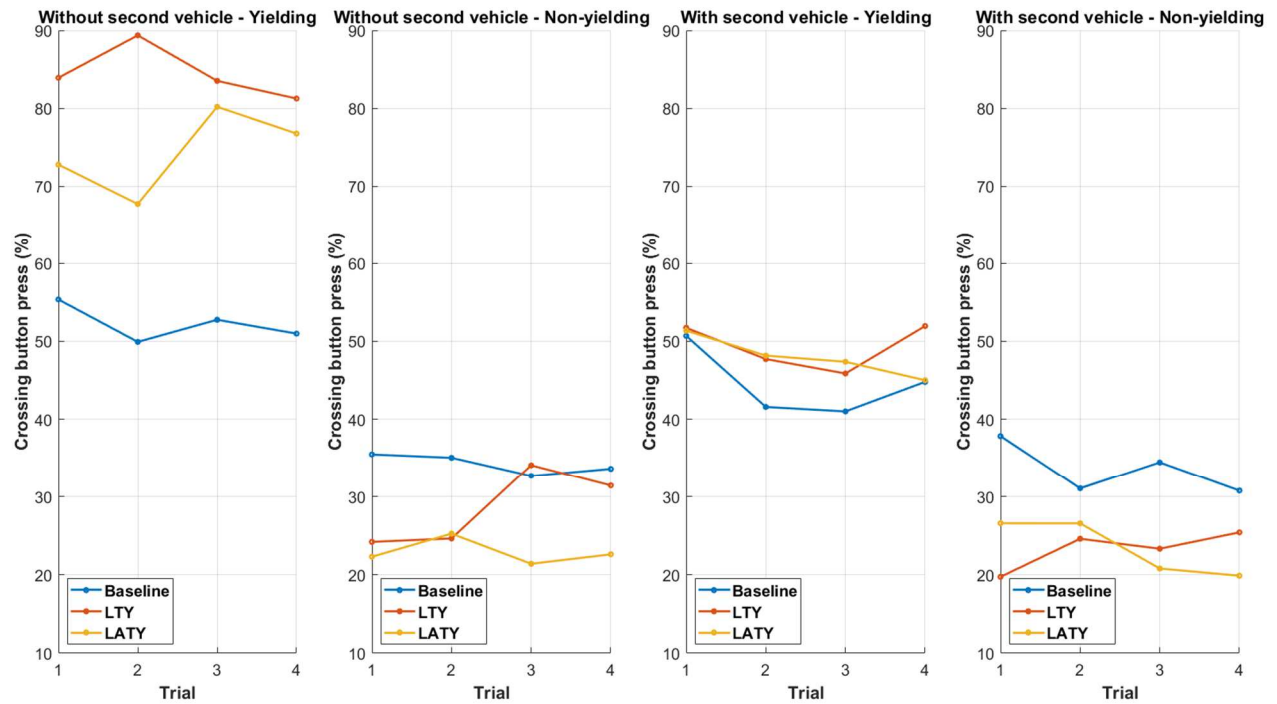


Figure S1. Mean crossing button press percentage as a function of trial number. The three mappings were presented in blocks with the Baseline block first, followed by the LTY or LATY blocks in counterbalanced order.